2016

Development of a Statistical Theory-Based Capital Cost Estimating Methodology for Light Rail Transit Corridor Evaluation Under Varying Alignment Characteristics

Anthony J. Catalina
University of Kentucky, ajcat@twc.com
Digital Object Identifier: https://doi.org/10.13023/ETD.2016.406

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Anthony J. Catalina, Student

Dr. Nikiforos Stamatiadis, Major Professor

Dr. Yi-Tin Wang, Director of Graduate Studies

DEVELOPMENT OF A STATISTICAL THEORY-BASED CAPITAL COST ESTIMATING METHODOLOGY FOR LIGHT RAIL TRANSIT CORRIDOR EVALUATION UNDER VARYING ALIGNMENT CHARACTERISTICS

DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Engineering at the University of Kentucky

By

Anthony J. Catalina
Louisville, Kentucky

Co-Chairs: Dr. Nikiforos Stamatiadis, Professor of Civil Engineering and Dr. Mei Chen, Professor of Civil Engineering
Lexington, Kentucky
2016

ABSTRACT OF DISSERTATION

DEVELOPMENT OF A STATISTICAL THEORY-BASED CAPITAL COST ESTIMATING METHODOLOGY FOR LIGHT RAIL TRANSIT CORRIDOR EVALUATION UNDER VARYING ALIGNMENT CHARACTERISTICS

The context of this research is the investigation and application of an approach to develop an effective evaluation methodology for establishing the investment worthiness of a range of potential Light Rail Transit (LRT) major system improvements (alternatives). Central to addressing mobility needs in a corridor is the ability to estimate capital costs at the planning level through a reliable and replicable methodology. This research extends the present state of practice that relies primarily on either cost averages (by review of cost data of implemented LRT projects) or cost categories in high and low cost ranges. The current methodologies often cannot produce accurate estimates due to lack of engineering data at the planning level of project development. This research strives to improve current practice by developing a prediction model for the system costs based on specific project alignment characteristics.

The review of the literature reflects a wide range of estimates of capital cost within each of the contemporary mass transit modes. The primary problem addressed in this research is the challenge associated with producing capital cost estimates at the planning level for the LRT mode of public transportation in the study corridor. Furthermore, the capital cost estimates for each mode of public transportation under consideration must be sensitive to a range of independent variables, such as vertical and horizontal alignment characteristics, environmentally sensitive areas, urban design and other unique cost-controlling factors. The current available methodologies for estimating capital cost at the planning level, by transit mode for alternative alignments, have limitations. The focus of this research is the development of a statistical theory-based, capital cost-estimating methodology for use at the planning level for transit system evaluations. Model development activities include sample size selection, model framework and selection, and model development and testing. The developed model utilizes statistical theory to enhance the quality of capital cost estimation for LRT investments by varying alignment characteristics.

This research has identified that alignment guideway length and station elements (by grade type) are the best predictors of LRT cost per mile at the planning level of project development. For the purpose of validating the regression model developed for this research, one LRT system was removed from the data set and run through the final multiple linear regression model equation to assess the model's predictive accuracy. Comparing the model's estimated cost to the projects final construction cost resulted in a 26.9% error. The percentage error seems somewhat high but acceptable at the planning level, since a 30% contingency (or higher) is typically applied to early level cost estimates.

Additionally, a comparison was made for all LRT systems used in the model estimation and the percent error range is from 2.4% to 111.5% with just over 60% of the project's predicted cost estimate within 30% or better. The model appears to be a useful tool for estimating LRT cost per mile at the planning level when only limited alignment data is available. However, further development of improved predictive models will become possible when additional LRT system data becomes available.

KEYWORDS: Planning Level Capital Cost Estimates, Light Rail Transit, Statistical Cost Estimating Models, Transit Corridor Planning and Evaluation, Transit Alignment Characteristics

Anthony J. Catalina

September 22, 2016

DEVELOPMENT OF A STATISTICAL THEORY-BASED CAPITAL COST ESTIMATING METHODOLOGY FOR LIGHT RAIL TRANSIT CORRIDOR EVALUATION UNDER VARYING ALIGNMENT CHARACTERISTICS

By

Anthony J. Catalina

Dr. Nikiforos Stamatiadis
Co-Chair of Dissertation

Dr. Mei Chen
Co-Chair of Dissertation

Dr. Yi-Tin Wang
Director of Graduate Studies

September 22, 2016

To the memory of my father, Dr. Timothy M. Catalina. I miss him every day!

To the memory of Dr. Charles C. Schimpeler, my great friend and professional mentor.

ACKNOWLEDGMENTS

I would like to express great gratitude to my committee chair, Dr. Nick Stamatiadis, who played a very important role in improving this research and who did not give up on me throughout several stops and starts. Without his guidance this dissertation would not have been possible. His assistance and challenges were most influential to completing my dissertation.

I would like to acknowledge and thank Dr. Mei Chen for serving as co-chair of my dissertation committee. Her support and advice throughout the dissertation process are greatly appreciated. I would like to thank committee member Dr. Reg Souleyrette for his friendly guidance and thought-provoking suggestions. I would like to express my sincere gratitude to committee member Dr. Arnold Stromberg, whose statistical expertise and guidance during the model development and data analysis/interpretation were invaluable. I would also like to thank Dr. Dwayne Edwards, Dr. Kamyar Mahboub and Dr. Karl Raitz for their participation in and support of my dissertation.

Finally, during the long journey to completing my dissertation, I have received tremendous encouragement from a great number of family and friends. My PhD would not have been possible without my family’s love and support. Special thanks to my mother, Theresa Balboa, and my two sons, Curtiss and Kai Catalina, who help me put everything into perspective and inspired me to complete this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION, PROBLEM STATEMENT, AND APPROACH

1.1 Introduction

Under the National Environmental Policy Act (NEPA) of 1969, every area seeking federal funding for a major capital transportation project must prepare a statement of the Purpose and Need of the transportation investments. Similarly, every area seeking federal funding for transportation improvements must have an ongoing planning process that includes community outreach and associated development of goals and objectives unique to that community. These two important inputs (Purpose and Need, and Goals and Objectives) will therefore exist for every major community project and form the basis for evaluation of the cost-effectiveness of alternative transportation improvement programs. This research advances the present state of practice at the system planning level for estimating the cost component of traditional cost-benefit analysis for Light Rail Transit (LRT) projects.

This dissertation investigates the development and application of an effective evaluation methodology for establishing the investment worthiness of a range of potential major transportation system improvements (alternatives) to address mobility needs with emphasis on the development of a reliable, replicable methodology for estimating LRT capital costs. The review of the literature reflects a wide range of capital cost estimates for the LRT mode of public transportation. Most of these estimates are based on a review of the cost data of implemented LRT projects; these estimates lack any systematic approach that can be followed in future evaluations of alternatives.

The primary problem addressed in this research is associated with the production of capital cost estimates at the planning level while considering LRT as an alternative mode of public transportation in the study corridor. Furthermore, the capital cost estimates for each mode of public transportation under study must be sensitive to a range of independent variables, such as vertical and horizontal alignment characteristics, environmentally sensitive areas, urban design, and other unique cost-controlling factors. Currently, deficient methodologies are used for estimating capital cost by transit mode for alternative vertical and horizontal alignments for a given corridor in which System Planning and Alternative Analysis (AA) are being carried out. These current methodologies rely on either LRT cost averages or a range of high and low cost categories, and are established without any consideration of the specific project alignment characteristics. Therefore, current approaches cannot produce accurate estimates due to a lack of engineering data at the planning level of project development.

Development of the methodology for evaluating transportation options must consider the purpose and need of the corridor for transportation improvements and the stakeholder/community goals and objectives for mobility and quality of life along the corridor. The appropriate evaluation process should include all potentially viable transportation options that meet purpose and need, as well as the goals and objectives for that community. The cost element of the proposed project plays an extremely important role in the decision-making process. The focus of this research is on the development of a statistical, theory-based capital cost-estimating methodology for use at the planning level in the cost effectiveness evaluation for analyzing LRT as an alternative transit system.

1.2 Problem Statement

The primary problem addressed in this research is the challenge associated with producing capital cost estimates for the LRT mode of public transportation in the study corridor, particularly during the early stages of project planning and development. The capital cost estimates for each mode of public transportation being considered as a candidate mode must be sensitive to a range of independent variables, such as (a) vertical alignment characteristics (the extent of the alignment that will be at-grade, elevated on fill, elevated on structure, in bored tunnel, in cut-and-cover tunnel, or in open trench); (b) horizontal alignment characteristics (right-of-way availability, land use types and densities, displacements and acquisitions); (c) environmentally sensitive areas (including wetlands, parks and open space, schools, hospitals, cemeteries); and (d) urban design (pedestrian walk-ways, access treatments, landscaping, station amenities) and other unique cost controlling factors (transit-oriented development (TOD) elements and station joint development features, etc.)

In almost all transportation improvement projects, every major development and implementation decision is politically motivated. The alternative evaluation methodology needs to assess how well potential transportation improvements address corridor mobility problems and needs to highlight the strengths and weaknesses of each option for decision-makers. The evaluation methodology should provide the basis for an objective assessment and a more informed decision-making process, even though it is often a very political and emotional decision at the local level. Cost typically plays a major role in the determination.

The cost associated with transportation improvements can be analyzed in great detail, both on the capital side and on the operation and maintenance side. Nothing is cheap; however, some modes involve massive capital investments while others consume large amounts of resources to run services and maintain hardware (Grava, 2003). The costs -- either in a project's entirety or by separate components -- are frequently, as might be expected, life-or-death factors for any transportation project (Grava, 2003). Accurate early cost estimates for LRT systems become critical for many parties, including project owners, when making investment decisions (Gunduz, Ugur, & Ozturk, 2011). More reliable capital cost estimates require a greater level of detail regarding LRT project design and construction. While the more reliable cost estimates are typically produced during the final design and pre-construction phases of project development, evidence from the current literature and industry practice reveals insufficient methods for developing reliable estimates at the planning level. During planning or feasibility studies, it is difficult to develop reliable cost estimates within a limited time frame when it is not possible to produce a detailed design. Consequently, other fast and accurate methods are required (Gunduz et al., 2011). This dissertation research emphasizes the importance of developing a methodology for more consistency in estimates of systems planning-level LRT capital costs.

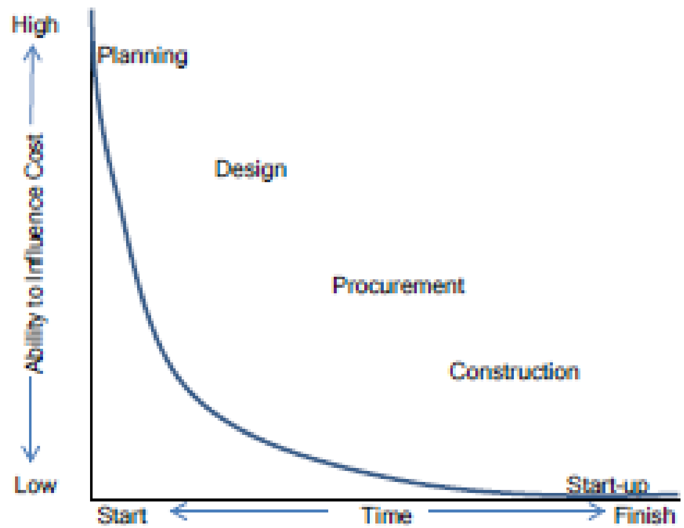
1.2.1 *Role of Capital Cost in Project Evaluation*

This research develops an approach to a rational and replicable capital cost estimating methodology to be used in cost-effectiveness evaluation for transportation systems considering the LRT mode. Although reliable methodologies for estimating benefits in evaluating transportation options for a specific purpose and need of a corridor for transportation improvements exist, methodologies for estimating capital cost for use in cost-benefit evaluation at the planning level are unreliable under the current state of practice. This lack of reliability is evidenced by the fact that if ten experienced transportation planners were asked to provide an estimate of the cost per mile of a given mode of transit, specifically for LRT, they would likely provide ten vastly different estimates. Furthermore, the aforementioned lack of reliability points to the deficiency in current cost-effectiveness evaluation methodologies, wherein estimates of effectiveness (or benefits) are derived from the application of reliable procedures while estimates of the cost component of the cost-effectiveness equation have no similar sophisticated estimating basis at the planning level.

The need for a more reliable methodology for planning-level capital cost estimates is underscored by the fact that at the planning level, engineering decisions can be made that significantly affect project costs. As project development advances from Systems and Conceptual Planning through Alternatives Analysis, Preliminary Engineering, and the various stages of Final Design into Construction, the ability to influence capital cost diminishes drastically. The relationship of planning level cost’s ability to influence transit capital costs through the stages of project development is presented in Figure 1.1 below.

In the cost-benefit equation, an accurate estimation of cost is equally important in optimizing the cost-effectiveness of an investment decision, as well as optimizing the determination of benefits. In addition, inaccurate estimations of LRT capital cost at the planning level can easily lead to erroneous conclusions regarding the investment worthiness or cost-effectiveness of the transportation options under study. When one recognizes that the greatest impact on the final cost of the constructed facility is possible at the planning level, one realizes the extent to which the accuracy of cost determination during the planning phase is critical.

Figure 1.1: Project Development Phases – Influence on Capital Cost



1.2.2 Transit Mode Analysis and Selection for this Research

Currently, methodologies for estimating capital cost by transit mode for alternative vertical and horizontal alignments for a given corridor are available but deficient. Most of these methodologies are based on a review of cost data of implemented LRT projects and lack any systematic approach that can be followed in future evaluations of alternatives. Completing the planning phases of the major capital investment evaluation process with inaccurate capital cost determination can easily lead to moving the project forward with a less than optimal modal candidate. In response, this research develops a statistically based methodology for estimating capital costs (based on alignment characteristics in the study corridor) for the most frequently encountered LRT mode of fixed guideway transit system development.

The objective of this research is to develop a capital cost estimating methodology for the LRT mode that can be replicated reliably in a full range of urban environments. This research identifies variables of concern that can be predictors of cost at the planning level, and focuses on capital cost estimating modeling for application and testing of the procedures in a complex urban setting.

1.3 Research Approach

A comprehensive literature review was undertaken to summarize the state-of-the-art in transit project evaluation with an emphasis on capital cost estimating, along with methodologies employed and limitations encountered in current practice. Other professional disciplines, including the Social Sciences and Medicine, have long used statistical theory in the development and evaluation of research results, as well as in the advancement of the state of practice. In this research, statistical theory (including regression analysis, analysis of interdependence, analysis of dependence, and the graphic presentation of statistical results) is studied regarding applicability in addressing the deficiencies identified above to enhance the reliability of developing cost estimates used for improving alternative transportation system development decisions.

A wealth of data has been gathered defining the basis for engineering decision-making at the System Planning level. Similar data exists for constructed transit projects at other important program development milestones -- AA, Preliminary Engineering (PE), Final Design (FD), and Construction. This research identifies the effective determinants of project cost for LRT projects and suggests the use of these determinants at the earliest level of system planning. An important source of this data is the information required by the Federal Transit Administration (FTA) and the NEPA-mandated major capital project development process.

Model development activities include sample size selection, model framework and selection, and model development and testing. The developed model utilizes statistical theory to enhance the quality of capital cost estimation for major transit investments by varying alignment characteristics.

Necessary statistical validations of cost estimates prepared by the models developed herein are undertaken. FTA data available at the Systems Planning and/or AA level are compared to actual cost at project completion, yielding a determination of the model's ability to predict costs successfully. The model developed herein can be embedded in the current cost-effectiveness evaluation techniques used today.

1.4 Organization of Dissertation

The purpose of this dissertation is to document the research efforts regarding the development of reliable capital cost prediction models that can be used as part of the evaluation methodology for alternative multimodal transportation systems. The research is directed towards the development of a reliable, replicable methodology for estimating LRT capital cost at the planning level as a crucial component in the overall evaluation methodology for transportation system improvements (alternatives) to address mobility needs in a corridor.

This dissertation documents analyses and findings of the research. The dissertation is organized as follows:

- Chapter 1 provides the transportation improvements evaluation background, problem statement, research approach, and organization of the dissertation.

- Chapter 2 summarizes the literature review regarding transportation improvements project evaluation, capital cost estimating methodologies, statistical analysis techniques regarding cost estimating, recent openings and planned construction for transit projects, and the FTA's capital cost estimating process.

- Chapter 3 presents the research methodology, data collection plan and sources, project selections and sample size, and statistical analysis methods.

- Chapter 4 introduces the model framework and selection, data preparation, and model development and testing, summarizes the statistical model cost estimating results, and presents the statistical validation of cost estimating results.

- Chapter 5 provides conclusions, the contribution of this research to the state of the practice, and recommendations for further steps in advancing this research.

2. LITERATURE REVIEW

The first task of this research was to perform an extensive literature review regarding the current state-of-the-art methodologies for evaluating alternative transportation improvements, specifically transit modes and LRT, and the development of capital cost estimates at the planning-level. The review was designed to identify recent literature as it relates to the research approach and problem statement in the following categories:

- Transit project evaluation methodology and supporting U.S. Federal guidance and regulations to support funding major capital transportation improvements
- Capital cost estimating methodologies and statistical analysis techniques for developing capital cost estimates
- Predicted vs actual cost of major transit projects currently in operation
- Recent openings and planned construction for transit projects
- Other related literature regarding the current state of the practice of transportation decision-making and cost estimate development, including the FTA capital cost estimating process

A thorough literature review was conducted using U.S. Department of Transportation [FTA and Federal Highway Administration (FHWA)] and US Environmental Protection Agency (EPA) sources and websites; Transportation Research International Documentation (TRID), an integrated database that combines the records from the Transportation Research Information Services (TRIS) Database and the OECD's Joint Transport Research Centre's International Transport Research Documentation (ITRD) Database; Ei Compendex (an engineering literature database); and other professional journals and websites, as well as public transportation agencies' websites.

The results and brief summaries of the literature review are provided below.

2.1 Transit Project Evaluations – State-of-the-Art Review

Multi-year investment planning is one of the most critical challenges for public transportation agencies and organizations. Competing transportation needs and initiatives -- combined with limited resources and funding issues -- can complicate the planning process. There are many competing objectives that must be reconciled in capital investment planning, including short-term vs. long-range investments, geographical vs. population-based project distribution, constituent demands vs. legislative requirements, and emergency repairs vs. preventative maintenance (Decision Lens, 2013).

Development of the methodology for evaluating transportation options must consider the corridor's purpose and need for transportation improvements. Additionally, this development must consider stakeholder/community goals as well as objectives for mobility and quality of life in the corridor. The evaluation process should include all potentially viable transportation options that meet the purposes, needs, goals, and objectives for that community.

The purpose of this review was to frame the overall transit alternative development and evaluation process where planning level capital cost estimates play an important role in the decision-making process. This literature review included the development and evaluation of public transportation improvement alternatives methodology and required U.S. Federal guidance and regulations to support funding major capital transportation improvements. These areas include:

- The transportation planning process
- Purpose and Need development
- Environmental analysis and review
- Public involvement process
- Alternative evaluation methodology
- Steps in following federal guidance and regulations for major capital investment decisions, approvals and funding

Cost is an extremely important element in the transit project evaluation and decision-making process. The development of a reliable methodology for the prediction of planning level LRT systems cost is needed at every stage of the evaluation process. This review is intended to present a summary of the complete picture of the transit evaluation sequence where the lack of accurate cost estimates can become an issue. Better methods for estimating cost are needed for inclusion in the overall transit alternative development and evaluation process where planning level capital cost estimates play an important role in the decision-making process. The background information pertaining to transit project evaluation is summarized in the following sections.

2.1.1 Transportation Planning Process

The development and evaluation of major capital transportation improvements (e.g., FTA New Start Projects), like all transportation investments in metropolitan areas, must emerge from a regional, multimodal transportation planning process in order to be eligible for Federal funding. Transportation planning includes a comprehensive consideration of alternative improvement strategies, an evaluation process, the collaborative participation of relevant public agencies, and meaningful public involvement. In urbanized areas, the transportation planning process is conducted by a Metropolitan Planning Organization (MPO) (FTA, 2016).

According to the Transportation Planning Capacity Building Program (TPCB) website:

> "In order for transit to play a meaningful role in creating sustainable transportation systems, transit agencies have to become more meaningful and pro-active partners in the Federal transportation planning and programming process. In order to get transit at the table (i.e., included in plans, funded, and built), transit agencies must first make their way to the table and actively participate in transportation decision-making. Recognizing that this may not be a traditional role for many transit agencies, the FTA sponsored a series of publications, supported by peer events and roundtables, called "Transit at the Table" that identifies ways for transit agencies to effectively participate in metropolitan and statewide transportation planning." (FHWA/FTA, 2016)

Some of these activities include: participating on the MPO Board and/or committees; collaborating with the business community, citizen groups, local officials, and other MPO partners; promoting land use/economic development/transportation integration, leading to MPO policy support for transit-oriented development; promoting early, open, and objective consideration of transit in regional corridor studies conducted by, or through, the MPO; assuming joint sponsorship of studies with state DOTs; and participating in preparation of the long range transportation plan and short range transportation improvement program (TIP), among others.

2.1.2 Purpose and Need Development

A well-conceived statement of the transportation problems for which potential options are being analyzed is a key early step in the corridor planning process. A Purpose and Need Statement is specifically required by the NEPA as a fundamental requirement when developing a transportation proposal that may require future NEPA documentation, i.e., an Environmental Impact Statement (EIS) or Environmental Assessment (EA). A concise, direct Purpose and Need Statement can help guide the development of any corridor level analysis. Purpose and need establishes the problems that must be addressed. Additionally, the statement serves as the basis for the development of goals and objectives, and provides the framework for determining which alternatives should be considered as reasonable options. Furthermore, some other federal transportation processes also require the generation of a Purpose and Need Statement in order to move a project forward in the project development process or to apply for funding. A Purpose and Need Statement will be required for any specific action for which NEPA review is required, such as any highway project, any public transportation capital project, and any multimodal project that requires an approval from FHWA or FTA.

The fundamental legal guidance on Purpose and Need Statements comes from the NEPA Council on Environmental Quality (CEQ) regulation, Section 1502.1: the Purpose and Need Statement "shall briefly specify the underlying purpose and need to which the agency is responding in proposing the alternatives including the proposed action." If it is anticipated that federal actions and federal funding will likely be a part of the implementation plan for mobility improvements in a corridor, the purpose and need can serve as the foundation for a specific action for which NEPA review is required.

The Purpose and Need Statement will provide the basis for consideration of transportation alternatives to improve mobility in a study and for decision-making and recommended actions identified that are carried forward for further study. The need identifies and describes underlying corridor mobility problems or deficiencies to be solved and provides supporting facts and analysis. The purpose is made up of a set of one or more goals and quantified, measurable objectives that must be met to a degree that sufficiently fulfills the underlying need. Each transportation alternative developed to address a corridor's Purpose and Need will have an associated capital cost. An enhanced planning level cost estimating methodology can be used to better define the cost and benefits of alternatives under consideration based on the Purpose and Need.

2.1.3 Environmental Analysis and Review

The NEPA establishes national environmental policy and goals for the protection, maintenance, and enhancement of the environment, and NEPA was signed into law on January 1, 1970. Essentially, NEPA "requires federal agencies to consider the potential environmental consequences of their proposals, to consult with other interested agencies, to document the analysis, and to make this information available to the public for comment before the implementation of the proposals" (Center for Environmental Excellence by AASHTO: NEPA Process). To address NEPA responsibilities established by the Council on Environmental Quality (CEQ), the FHWA and the FTA issued a set of regulations titled (23 CFR § 771) *Environmental Impact and Related Procedures* (FHWA/FTA, 2016).

Depending on the potential environmental impacts of the transportation project, three environmental clearance options are available under NEPA. The required documentation depends on whether a project significantly affects the environment. These three options are:

- **Categorical Exclusions (CE) -** Projects that do not individually or cumulatively have significant environmental effects are classified as categorical exclusions

- **Environmental Assessments (EA) -** If the significance of environmental impacts is unknown, preparation of EA typically is required to determine whether an EIS is necessary. If no, the agency issues a finding of no significant impact (FONSI). If yes, an EIS will then be prepared.

- **Environmental Impact Statements (EIS)** – Under NEPA, EISs are required when there is a proposal for a major federal action that significantly affects the quality of the human environment. An EIS includes a detailed evaluation of the proposed action and alternatives. The public, other federal agencies and outside parties may provide input into the preparation of an EIS and then comment on the draft EIS. It should be noted that most major capital transportation improvements involving fixed guideway transit, such as the subject of this research (LRT capital cost estimates), require an EIS.

The proposed corridor alignment for an LRT alternative may transverse near or through environmentally sensitive areas and impact the environment (natural, physical, or human) in other direct or indirect ways. Some of the sensitive areas include wetlands, parks and open space, schools, hospitals, and cemeteries, among others. The avoidance and/or mitigation of potential impacts have associated cost that is included in the project's capital cost estimates. Although difficult to quantify at the planning level, LRT alignment characteristics that have potential environmental impacts should be considered in the development of early level cost estimates.

2.1.4 Public Involvement Process

All major transportation improvement projects will include extensive public involvement. "Public participation is an integral part of the transportation planning process, which helps to ensure that decisions are made in consideration of and to benefit public needs and preferences. Early and continuous public involvement is a key input to the transportation decision-making process. Successful public participation is a continuous process, consisting of a series of activities and actions to both inform the public and stakeholders, and to obtain input from them which influences decisions that affect their lives" (FHWA, 2016). According to the FHWA website:

> "Federal statutes and regulations provide general guidelines for public involvement. Transportation agencies have great flexibility in developing

> specific public involvement and public participation plans. Techniques for each situation may differ, depending on things like demographics and specific impacts of a project, but the general approach to developing a public involvement and public participation plan should contain elements that are relevant to all communities." (FHWA, 2016)

The federal regulations and guidelines regarding public involvement provide great flexibility on how to conduct effective outreach while also emphasizing its importance to the transportation planning process.

The capital cost estimate of an LRT project includes costs for all planning, design and construction. The cost of the public involvement program is included in the final capital cost. For many transit projects, the capital cost estimate becomes a point of contention among the public participates, leading to project delays and requiring more extensive outreach, thereby increasing the cost of the public involvement process. It is important to have the ability to develop accurate planning level capital cost and to clearly present the cost estimate development process to the public.

2.1.5 Alternative Development and Evaluation

The overall approach of the alternative development and evaluation process involves addressing what specifically needs to be done to improve corridor mobility, and explains exactly how to accomplish this goal using the wealth of information generated through the transportation planning process. The purpose is to develop and evaluate a range of multi-modal potential transportation improvements that address key mobility issues in the study corridor. Potential options and service delivery strategies are typically drawn from the local agency's Long Range Transportation Plan (LRTP) or Regional Transportation Plan (RTP), and new options generated by addressing the purpose and need.

A larger universe of potential options should be narrowed down to a more manageable list, which sharply focuses on corridor mobility issues and travel markets. All options studied must have independent utility; that said, they may also involve potential packaging of options into systems or alternatives to create a comprehensive corridor mobility improvement. The second screening level should narrow the candidate options to a smaller number that is ultimately expected to lead to a selection of a locally preferred alternative (LPA), which will be taken forward from the AA phase through the major project development process (PE, EA/EIS, and FD).

Alternatives analysis should clearly indicate why and how the particular range of project options was developed. All reasonable alternatives should be considered and discussed at a comparable level of detail, including the No-Build Alternative and Transportation System Management alternatives. Alternatives analysis should explain why and how alternatives were eliminated from consideration, documenting the criteria used to eliminate alternatives as well as the measures for assessing the alternatives' effectiveness.

The purpose of the evaluation methodology for the alternatives is to assess how well potential transportation system improvement options address corridor mobility problems, and to highlight for decision-makers the strengths and weaknesses of each option. Typically, a two-stage evaluation process is developed in which increasingly detailed and comprehensive performance measures are applied to a decreasing number of options identified as the best potential transportation system improvements. Both quantitative and qualitative evaluation criteria should be used to differentiate between options and ensure that each option considered has independent utility while addressing key corridor mobility issues as well as local preferences/values. Cost is almost always one of the quantitative criteria used the evaluation methodology for LRT projects. While the more reliable cost estimates are typically produced during the final design and pre-construction phases of project development, evidence from the current literature and industry practice reveals insufficient methods for developing reliable estimates at the planning level. This research emphasizes the importance of developing a methodology for more reliable estimates of systems planning-level LRT capital costs.

2.1.6 Federal Guidance and Regulations – Capital Investment Grant Program Evaluation and Rating Process

Per the FTA’s Capital Investment Grant (CIG) program overview:

> “The U.S. Department of Transportation makes recommendations to the U.S. Congress regarding the justification and merits of funding competing U.S. major capital transportation projects. These recommendations are made pursuant to 49 U.S.C. Section 5309. The Federal Transit Administration’s (FTA) discretionary Capital Investment Grant (CIG) program provides funding for fixed guideway investments.” (FTA, 2016)

The categories of eligible projects under the CIG program are:

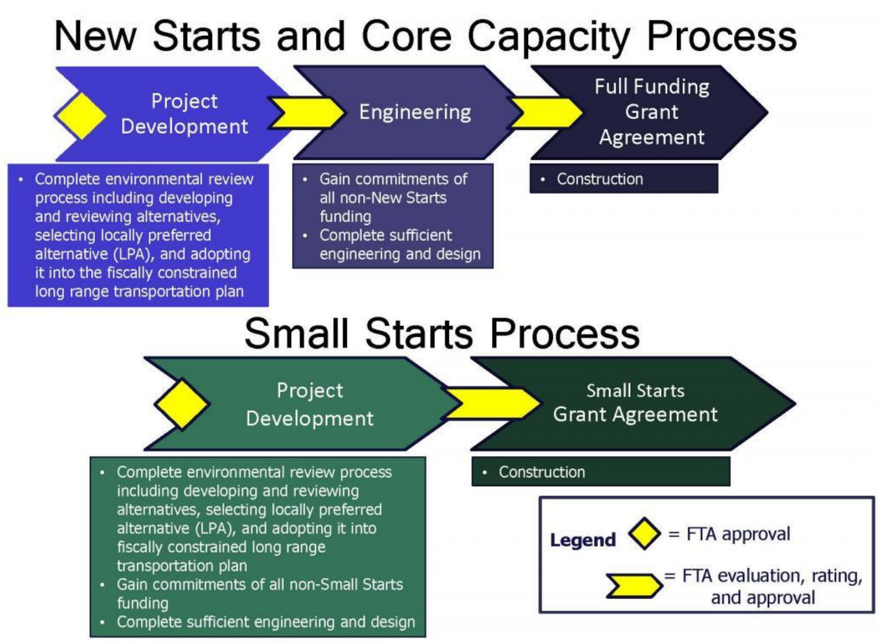
- New Starts – New fixed guideway projects (or extensions) with capital cost of $300 million or greater, or requesting $100 million or greater from FTA (CIG 5309 funds)

- Small Starts - New fixed guideway projects (or extensions) to systems or BRT projects with capital cost of less than $300 million or requesting less than $100 million from FTA

- Core Capacity - Investments in existing fixed guideway systems that increase capacity by at least 10 percent in corridors that are currently at capacity or will be in the short-term future (within five years)

The law requires a specific project development process and timeframes to be eligible for CIG funding. New Starts and Core Capacity require completion of two phases in advance of receipt of a construction grant agreement – Project Development and Engineering. Small Starts projects are required the completion of one phase in advance of receipt of a construction grant agreement – Project Development. The law also requires projects to be rated by the FTA at various points in the process according to statutory criteria, evaluating project justification and local financial commitment (FTA, 2016). The capital cost estimate for the transit project is a critical element used in computing both the project justification and local financial commitment ratings. This emphasizes the importance of developing a methodology for more reliable estimates of systems planning-level LRT capital costs, specifically used during the early project rating stages during Pre-Project Development, when detailed engineering data has yet to be developed. The FTA project development process is represented below in Figure 2.1.

Figure 2.1: FTA Project Development Process

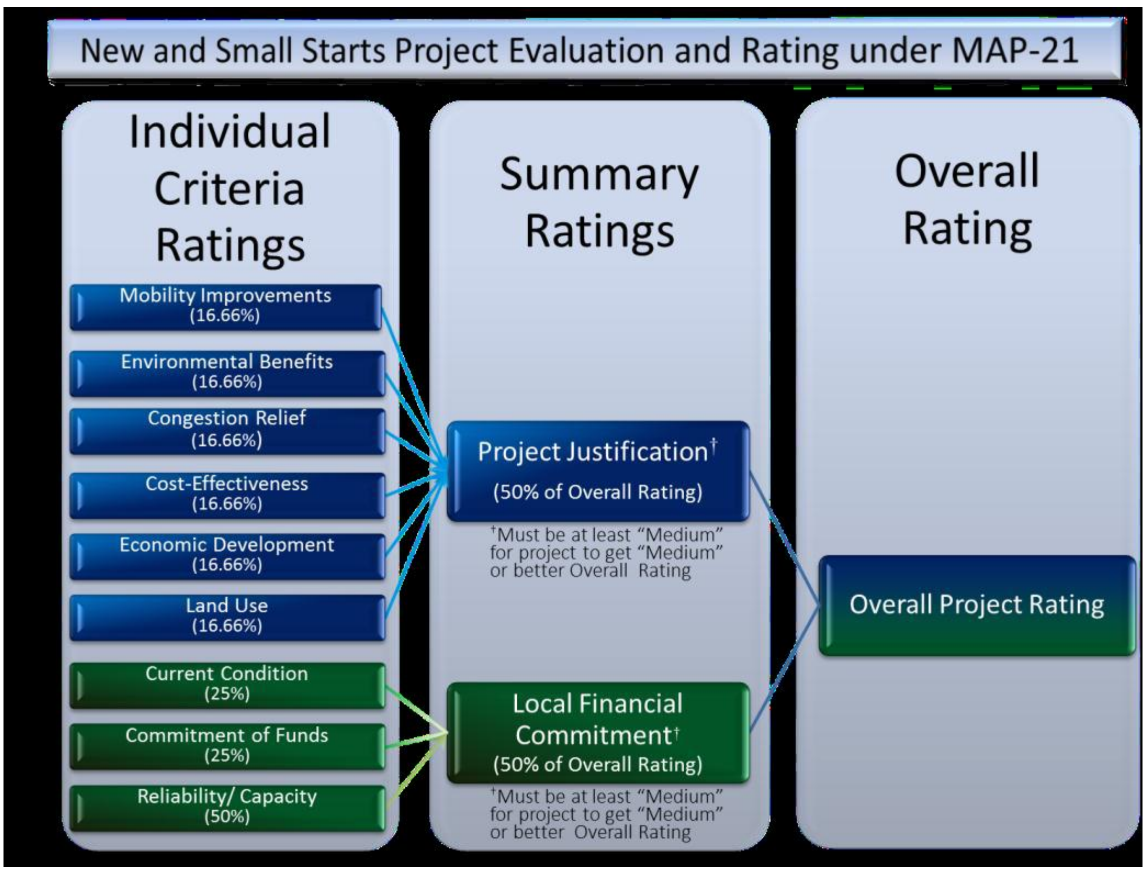


Source: FTA

The CIG program outlined in 49 USC 5309 was most recently authorized in December 2015 by the Fixing America’s Surface Transportation Act (FAST). The CIG program is the federal government’s primary financial resource for supporting transit capital projects that are locally planned, implemented, and operated. It provides funding for fixed guideway investments, such as new and expanded heavy rail, commuter rail, light rail, streetcars, bus rapid transit, and ferries, as well as corridor-based bus rapid transit investments that emulate the features of rail (FTA, 2016).

The following chart (Figure 2.2) outlines the weights given to the various criteria and how they are combined into summary ratings and an overall rating. The transit project's capital cost is a key input to the Cost-Effectiveness criteria rating.

Figure 2.2: New and Small Starts Project Evaluation and Rating



Source: FTA

According to the Capital Investment Grant Program's *Annual Report on Funding Recommendations:*

> " FAST establishes various criteria on which proposed CIG projects must be evaluated and specifies a five-point rating scale: *High*, *Medium-High*, *Medium*, *Medium-Low*, and *Low*. To advance in the process toward a funding recommendation in the president's budget and a construction grant

agreement, a project must be rated *Medium* or higher overall. Receipt of project funding through a construction grant agreement is subject to congressional appropriation, and is only obligated when the grantee can assure the FTA that the proposed project scope, cost estimate, and budget are firm and reliable, local funding commitments are in place, and all critical third party agreements have been complete." (FTA, 2016).

2.2 Capital Cost Estimating – Current Methodologies and Limitations

The primary problem addressed in this research is the challenge associated with producing capital cost estimates at the planning level for the LRT mode of public transportation in the study corridor. The development of cost-estimating methods and models for transportation projects as well as statistical analysis techniques for developing estimates are summarized here.

2.2.1 Capital Costs

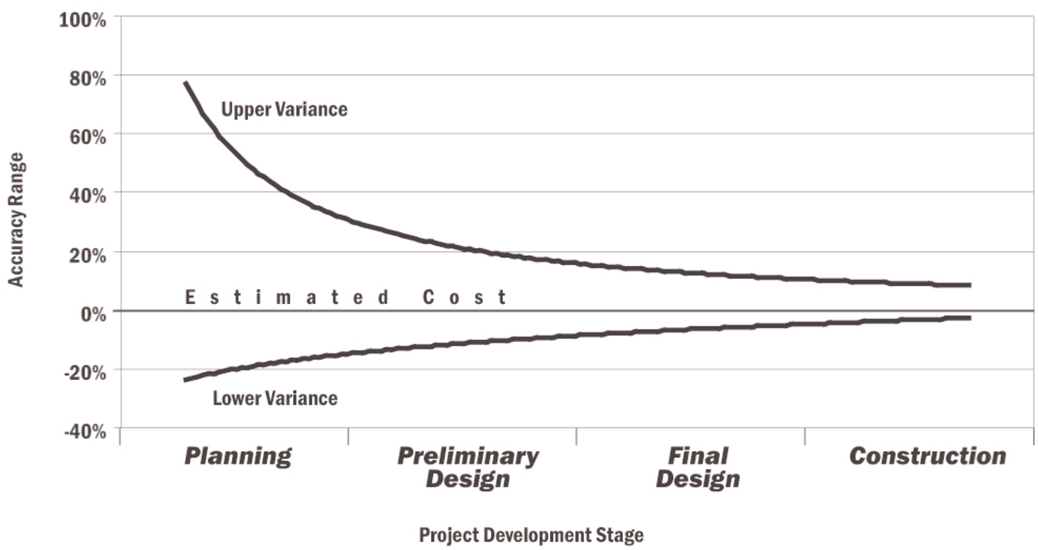
Capital cost estimates are an important element in calculating the cost-effectiveness, financial requirements and implementation feasibility of major capital transit investments. Reliable capital cost estimates are key to the evaluation process and the decision-making regarding preferred alternatives. An important element of the cost estimating methodology, especially at the planning level of analysis, is the development of unit costs, such as costs per mile, for capital elements of the transit system.

Capital cost estimates that are prepared for transportation projects during the planning phase, typically defined as being less than 30 percent completion of engineering design, are an important criterion used by decision-makers to evaluate the feasibility and cost effectiveness of proposed projects. In recent years, a number of high profile transportation projects have received media attention because of significant cost increases that occurred between the time the projects were planned and the time construction began. This apparent disparity has drawn the attention of elected officials and has resulted in efforts to improve the accuracy of the process (Harbuck, 2007).

According to Harbuck (2007), the accuracy of a cost estimate is generally defined as the closeness between the value predicted by a cost estimate and the final constructed value, typically expressed as a +/-percentage based on a defined scope of work. The accuracy of estimates typically tends to improve as the level of project definition advances. For transportation projects, the level of project definition is primarily a function of the level of engineering design completed to date. This variability in the accuracy of cost estimates based on the level of project definition is often shown graphically by the use of a "trumpet diagram," as shown in Figure 2.3 (Harbuck, 2007).

While Harbuck (2007) examined the accuracy of cost estimates for a range of transportation projects, he did not specifically study LRT at the planning level. By using the unit costs from several recently implemented LRT systems, this research aims to improve the accuracy and comparative analysis of LRT capital cost estimates at the planning level of a study. The importance of developing a more reliable methodology for planning level capital cost estimates is underscored by the fact that, at the planning level, engineering decisions can be made that significantly affect project costs. As project development advances from planning through design (Preliminary Engineering and Final Design), and into Construction, the ability to influence capital cost diminishes drastically, which ultimately impacts accuracy.

Figure 2.3: Estimating Accuracy Trumpet



Source: Harbuck (2007)

2.2.2 Cost Estimating Methods, Models and Statistical Techniques

In evaluating transit systems, cost estimating has always been a major consideration. Hsu (2012) found that without an applicable cost estimating model or methodology, the choice between LRT and Bus Rapid Transit (BRT) would cause significant controversy and difficulty in the early corridor planning stage. The model Hsu developed can be applied to estimate costs and compare the LRT and BRT systems that operate on various right-of-way categories, alignment configurations, and different transit demand volumes (Hsu, 2012). Hsu developed a cost-estimating model for LRT and BRT systems to better aid transportation planners and decision makers in the selection process. Hsu's model examined the unit costs allocated to the cost components of capital and operations and maintenance (O&M) costs. These costs in his estimating model were derived from high-low unit costs of the historical data of existing systems obtained from multiple resources.

Most previous cost models and system comparisons do not take both capital and O&M costs into consideration, while the model developed by Hsu (2012) does. As a simplified planning tool for evaluation and comparison, Hsu's cost-estimating model can be used for evaluation and comparison to help determine which system would be a better mode for the respective area, so long as the passenger demand volume is given (Hsu, 2012). Hsu's model included historical cost information data for five LRT systems provided by the FTA for projects implemented prior to 1992 that included the classification of the capital cost into eight cost components.

Hsu's (2012) model provides an effective planning tool for evaluating and comparing the LRT and BRT modes to distinguish which may be a more applicable system for a given study area. The model utilized high and low cost ranges of cost components to develop estimates. Hsu's model estimates capital cost, O&M cost, and cost per mile; it also requires the passenger demand volumes of the subject systems in order to develop these cost estimates. However, at the early stages of planning, the travel demand volumes are often not available. Therefore, the model to be developed here will not require the passenger demand volumes, unlike Hsu's model. This dissertation's approach to developing planning level cost estimates focuses only on the capital cost component to derive cost per mile. Rather than relying on high and low cost ranges, this research is based on the statistical relationship and significance of the response (dependent) variable (cost per mile) to predictor (independent) variables.

The focus of this dissertation expands Hsu's research by focusing on the development of a statistical regression model for predicting future LRT cost per mile (compared to high – low ranges), the inclusion of several additional potential predictor variables (44 compared to 12), and the addition of 22 more (27 compared to 5) of the most recently implemented LRT systems in the U.S. entering into revenue service from 1986 (Portland – MAX Segment I) to 2012 (Pittsburgh – North Shore LRT Connector).

Hoback (2008) analyzed LRT costs for 24 (1985 – 2005) LRT lines and extensions with a sensitivity analysis to find unit costs for types of right-of-way (ROW) construction. Hoback used a sensitivity analysis to limit the errors and arrive at the final unit costs for each ROW type. The purpose of his analysis was to develop a tool that can quickly produce a rough estimate of light rail construction based only on mileage because of the limited project information that is known during conceptual development.

Hoback (2008) stated that many cost-estimating methods are too complex for feasibility study cost estimation because details regarding stations, vehicles, systems, and other elements are not available. Hoback observed that during feasibility studies, the researcher finds a system similar to the one proposed and then scales the cost up or down. Hoback's method applied a greater detailed approach of comparison with existing systems to find the schematic costs of units, such as a mile of elevated track, to aid in developing more accurate feasibility studies. The costs of light rail lines and extensions were found by comparing 24 construction projects. Additionally, engineering estimation and sensitivity analysis were used to minimize the discrepancy between the cost estimates and actual costs. Hoback (2008) formulated a sensitivity analysis examining the variation in system costs (output) as a function of the cost of ROW types (inputs).

Hoback (2008) concluded that estimation of schematic costs per mile for each type of right-of-way provided a tool for conceptual planning of LRT systems. This dissertation builds on Hoback's (2008) research by including additional LRT cost categories (eight categories divided into 44 potential cost predictor variables) beyond ROW types, and additional LRT systems (27 systems -- several differing -- implementing revenue service between 1986 and 2012).

Models can be used to develop parametric estimates for transportation projects. Models can be described as simple composite costs on a per unit basis that are developed from a set of more complex cost parameters. The parametric estimating methodology can be defined as an estimating technique that uses validated cost-estimating relationships (CERs) to estimate cost. These CERs rely on known or proven relationships between item characteristics and the associated item cost. The cost relationships can range from simple to complex and may use cost-to-cost variables or cost-to-non-cost variables (Harbuck, 2002).

Harbuck (2002) outlined the basic steps needed for preparing parametric estimates for transportation projects by using models. The first step in the development of a parametric estimating methodology is the development of a cost database. The cost data forms can include item level unit costs and composite unit costs, and can be developed from several sources. Harbuck (2002) stated that cost data can also be compared to costs for similar types of construction seen in the project region. He identified that the second step is the development of composite section cost models for the various cost categories that are required for the type of transportation project being analyzed. Harbuck (2002) presented the example for a major capital transit investment which uses the FTA's Standard Cost Categories (SCC), which are used in this dissertation's cost model development and identified and defined for this dissertation in Chapter 3. Harbuck (2002) stated that the final step is the presentation of the cost estimating results format, which should facilitate responses to different questions regarding the results of the cost estimates.

Kouskoulas (1984) developed a methodology for estimating predesign cost estimation functions of transportation systems that emphasized the methodology and the evaluation of the estimation equation. "The use of preliminary cost functions for estimating the cost of transportation systems (corridors, highways, pavements, excavation) is extremely useful for predesign estimates, budgeting, economic studies of highway development, taxation, and investment strategies" (Kouskoulas, 1984). Kouskoulas showed that the predesign cost estimation functions developed for highway pavement and for mass transportation systems prove to be useful in estimating the cost of highway pavements and mass transportation systems at the stages of early project development. Steps in his methodology include identification of cost functions, evaluation/analysis of estimators, and error analysis. Kouskoulas (1984) observed that the variables that characterize a mass transit system are many. Furthermore, the proper selection of a few good ones, and the proper use of sample data, even when limited in size, may lead to an accurate estimator.

Kouskoulas's (1984) analysis included 10 samples from nine large cities for the transit system (heavy rail and other modes). He stated that for such a limited size of sample data, the proper selection of the variables is important. His analysis concluded that the conditions necessary for the determination of a reliable cost estimation function are: (1) the selection of the proper objective and subjective variables which adequately describe the cost of the transportation system; (2) the collection of sufficient historical data, its availability and constant updating; and (3) the proper testing of the regression equation as to its correlation with the cost, the propriety of its linearity, and the significance of the individual variables. He also showed the probability of the regression equation failing to specify the cost function it is intended for and that the correlation and the significance of each variable are equally important for the final definition of the cost function.

This dissertation builds on Kouskoulas's (1984) research by including several more subjects on the transit side (27 vs 9) and focusing specifically on the LRT mode. This research is also able to use much more recent and validated capital cost data, as the reporting systems and documentation required by the FTA have greatly advanced in recent years.

Gunduz , Ugur, & Ozturk (2011) developed statistical methods for parametric cost estimation for transit systems. Their research presented multivariable regression and artificial neural network approaches for development of early cost estimation models for light rail transit and metro track works. They stated that the model developed for railway superstructure was not dependent on the type of the section of the transit line (e.g. Tunnel Boring Machine tunnel, depressed open/close or grade line, etc.) Regardless of transit mode (LRT vs Heavy Rail Transit (HRT)) for developing early cost estimates, they used track work as the independent variables. The data for their research came from 16 urban rail projects (7 HRT and 9 LRT) in Turkey, and the track data were analyzed by means of parametric cost estimation models (regression and artificial neural networks). They observed from their literature review that if there is limited information regarding the relationships between dependent and independent variables, artificial neural networks performed better than regression models; if the relationship between different variables can be identified, regression techniques performed better.

Gunduz et al. (2011) formulated 17 variables that best describe the track work cost from their data set. The list of variables included: total length of main trackway, length of ballasted trackway, length of direct fixation trackway, number of crossovers, maximum slope of the line, maximum super-elevation, minimum horizontal curvature, minimum vertical curvature, and cost, among others. Based on the results of each cost-estimating technique, they concluded that regression analysis estimated the cost of the validation projects with an error of 2.32%, while the artificial neural network estimated the cost with an error of 5.76%. They established that both of these models can be beneficial in the early decision-making phase of projects that includes track works.

This dissertation has a similar focus, but differs from Gunduz et al. (2011) in several areas. They examined both multivariable regression and artificial neural network approaches and included only track work elements requiring great detail as the independent variables. They also included both LRT and HRT systems operating in Turkey as the observations for development of early cost estimates. This dissertation focuses only on statistical techniques (regression for example), considers multiple cost categories (not just track work) for the independent variables and, uses only LRT systems recently constructed in the U.S. as the observations. Building on the Gunduz et al. (2011) approach, the model developed in this dissertation can be used as a simplified planning tool during early project phases when much of the detail regarding track work would not be known.

Other advanced statistical analysis techniques to develop early level cost estimates have been researched. Early estimates are critical to the initial decision-making process for the construction of capital projects (Trost and Oberlender, 2003). Estimates developed during early stages of capital projects by various industry methods are typically plagued by limited scope definition and are often prepared under time constraints. Furthermore, reliable cost data is often difficult to obtain during the conceptual stages of a project (Trost and Oberlender, 2003).

According to Trost and Oberlender (2003), a contrast arises when comparing the importance of early estimates with the amount of information typically available during the preparation of an early estimate. Such limited scope definition often leads to questionable estimate accuracy. Even so, very few quantitative methods are available that enable estimators and business managers to objectively evaluate the accuracy of early estimates.

Trost and Oberlender (2003) conducted research and developed a quantitative method for this objective. To accomplish this objective, quantitative data was collected from completed construction projects in the process industry [industries where the primary production processes are either continuous or occur in a batch of materials that is indistinguishable (e.g., food, beverage, plastics, paper products)]. Each of the respondents was asked to assign a one-to-five rating for each of 45 potential drivers of estimate accuracy for a given estimate. The data was analyzed using factor analysis and multivariate regression analysis. The factor analysis was used to group the 45 elements into 11 orthogonal factors. Additionally, multivariate regression analysis was performed on the 11 factors to determine a suitable model for predicting estimate accuracy. The resulting model, known as the estimate score procedure, allowed the project team to score an estimate and then predict its accuracy based on the estimate score (Trost and Oberlender, 2003). The multivariate regression analysis identified 5 of the 11 factors that were significant at the 10% level. The five factors, in order of significance, were basic process design, team experience and cost information (which highlights the importance of human factors and experience level in preparing cost estimates), time allowed to prepare the estimate, site requirements, and bidding and labor climate (Trost and Oberlender, 2003).

Although this dissertation does not utilize these advanced statistical techniques, they are worth future investigation for producing accurate cost estimates at early level transportation project planning phases.

2.2.3 Limitations of Current Planning Level Cost-Estimating Methodologies

During review of the literature regarding methodologies for capital cost estimating at the early stages of transportation project development, some limitations were identified. Harbuck (2007) observed that one of the foundational conditions needed for obtaining better accuracy in cost estimating is a defined scope on which the estimate is based. He emphasized the importance of developing, documenting, and communicating the project scope because deficiencies create cost estimate inaccuracies. He cited that one reason for weaknesses in the scope definition is the lack of engineering information available at the planning level of project development.

Harbuck (2002) also describes limitations of estimates of capital cost during the conceptual phase of a project as the evaluation and treatment of uncertainty. He identified four potential sources of uncertainty, which include:

- Changes in project scope
- Changes in design standards
- Incorrect unit cost/quantity assumptions
- Unforeseen problems in implementation

Hsu (2012) established that the development of planning level cost models can be accomplished through the application of unit costs from existing systems, but that the unit cost variation in the historical data could be largely based on the existing system's characteristics, locations, and other uncertainties, which makes accurate unit values difficult to collect. Hsu identified some uncertainties as design modifications, environmental conditions, engineering conditions, construction schedules and price inflation.

Hoback (2008) stated that in conceptual project development, sometimes the only known information available for developing cost estimates is the approximate system mileage and potential right-of-way type. Lack of proposed system data, such as alignment characteristics and engineering detail, obviously limits the ability to produce reliable and accurate cost estimates during planning phases.

2.3 Predicted and Actual Cost of Major Transit Systems

The FTA has conducted three studies (1990, 2003 and 2007) that analyzed the predicted actual capital cost and ridership of major transit projects that were constructed utilizing federal funds.

In the most recent study (FTA, 2007), the FTA conducted an analysis of the predicted capital cost as compared to actual final construction cost of 21 recently opened major transit projects that were constructed using funds under the New Starts program.

The analysis had two main purposes (FTA, 2007):

- To provide an up-to-date assessment of the actual performance of projects compared to the forecasts made for those projects; and
- To consider the effectiveness of the procedures and technical methods used to develop information for decision-making by project sponsors and the FTA

According to the FTA 2007 Report:

> "The analysis of the predicted and actual impacts of New Starts projects focuses on the reliability of the planning information used to evaluate and select projects for funding. FTA based this analysis on an inventory of ridership forecasts and cost estimates prepared at various stages of the project planning and development process. The data sources included environmental documents, AA studies, Major Investment Studies (MIS), New Starts application submissions, Full Funding Grant Agreements (FFGA), and Project Management Oversight Contractor (PMOC) reports. This information was then compared to the actual results reported by the project sponsors for ridership and by the PMOCs for capital costs" (FTA, 2007).

Cost estimates developed during an AA are used to support the local decision to choose an LPA and are generally the cost estimates that are presented to the FTA when projects apply to begin preliminary engineering. The decision to adopt an LPA signifies that the local decision-makers have chosen the specific mode (highway, LRT, BRT, etc.) and general alignment of a project to address the identified problems and needs in a corridor (FTA, 2007). FTA considers this decision, made at the end of AA, to be the most critical decision in the planning and project development process because the LPA decision has more impact locally than any subsequent local decision. It also provides the entry point information on costs, benefits and funding in the federal New Starts process (FTA, 2007).

The reliability of early project phases (planning level) cost estimates are important because they are used by decision-makers to determine if a project should advance to the next level of project development.

A review of the three previous FTA studies presents a great variation among the projects regarding their forecasted and actual capital costs. The Pickrell Report (Pickrell, 1990) first published an analysis of the predicted and actual impacts of 10 major capital transit projects (four HRT, four LRT, and two Downtown People Movers). Capital cost estimates were inaccurate. Cost estimates of two projects were within 20 percent of the original cost estimate while seven of 10 projects were between 30 and 100 percent higher than their original estimates. The cost estimate of one project was more than double (over 100 percent of) its cost estimate (FTA, 2007).

The 2003 FTA study (FTA, 2003) included 19 additional projects (LRT, HRT, Commuter Rail Transit (CRT)) that had been completed between 1990 and 2002. The FTA found that cost estimates had improved since the Pickrell Report, but these estimates still underestimated actual costs (FTA, 2003).

The capital cost findings from the 2007 FTA study identified that for 21 projects completed between 2003 and 2007, on average, actual constructed costs were greater than estimated (inflated) costs developed in earlier project development phases. The table below (Table 2.1) summarizes both the 2003 and 2007 study findings of the New Start projects.

Table 2.1: As-Built Capital Cost Compared to Inflation – Adjusted AA, FD, FFGA Estimate

Project Phase Cost Estimate	2003 Study	2007 Study
Alternative Analysis (AA)	20.9 %	40.2 %
Final Design (FD)	13.5 %	11.8 %
Full Funding Grant Agreements (FFGA)	7.3%	6.2%

Source: FTA 2003 and FTA 2007

The actual capital costs compared to the originally predicted costs of the projects analyzed in the three studies show that the accuracy improved from the 1990 study to the 2003 study by approximately 30 percent on average, but then decreased by approximately 20 percent on average from the 2003 study compared to the 2007 study. The analyses showed a positive trend: that prediction models are improving but still greatly underestimating cost when compared to the final construction cost, demonstrating the need for development of methodologies to improve planning level LRT (as well as other modes) estimates. The actual capital costs compared to the originally predicted costs of the projects examined in the 1990, 2003 and 2007 studies are presented in Table 2.2.

Table 2.2: As-Built Capital Cost Compared to Inflation – Adjusted AA/DEIS Estimate

Study	Average	50 Percentile
Pickrell Report 1990	150 %	151 %
FTA 2003	121 %	115 %
FTA 2007	140 %	122 %

Source: FTA, 2008

Dantata et al. (2006) examined trends in U.S. rail project cost overruns by presenting comparisons of the results of the Pickrell Report to cost overruns (original planning cost estimates as compared to final costs) of transit projects completed after 1990. They compared those statistics with data from 16 recent (from 1994 on) transit rail projects, with the objective of examining if the results in the Pickrell Report are still valid or if the magnitude of cost overruns in rail transit projects has changed. The comparison was conducted at a macro level and did not identify the potential causes of overruns. They stated that several factors may cause the cost overruns, such as the optimistic underestimation of costs in conceptual phases, the lengthy project approval and construction processes, the omission of project components during early phases and the addition to project scope during project development, among other factors.

Dantata et al. (2006) observed that there is evidence to indicate that cost overruns for projects completed in the Pickrell Report (before 1990) differ from those of projects completed after 1994, but did not have sufficient data to statistically prove this at a level of significance of 5%. However, their analyses showed a positive trend that cost overruns are improving and suggested the subject of their research be continued as more transit projects are completed and more data becomes available.

A condition of receipt of federal New Starts funds is that the project sponsors conduct an evaluation called a Before-and-After Study (FTA, 2016). The studies are used to determine the cost and ridership impacts of the transit project. Upon completion, this study must be submitted to the FTA. The FTA requires Before-and-After Studies on all projects that receive FFGAs.

The two purposes of Before-and-After Studies are to document the actual outcomes of major capital projects and to monitor the accuracy of predicted outcomes in order to identify methods and procedures that might merit subsequent attention to improve their reliability. Federal transit law requires that sponsors of projects receiving funds through the FTA discretionary capital-grant program prepare a Before-and-After Study if the project is (FTA, 2016):

- A New Starts or Small Starts project developed under SAFETEA-LU procedures; or
- A New Starts or Core Capacity project developed under MAP-21 procedures.

2.4 Recently Opened Transit Projects and Planned Construction Starts

The review of the literature in this section documents the wide range of estimates of capital cost within each of the contemporary mass transit modes. These capital cost estimates within the LRT mode are in a database of major transit investments (2015) provided by (Freemark, 2016) *The Transport Politic* (Openings and Construction Starts Planned for 2016), which includes 30 LRT projects, range from $34M per mile to $928M per mile. These LRT projects cover approximately 210 miles at an estimated average cost per mile of $196 million. The wide cost variation is based on several factors, specifically the corridor alignment characteristics of each project. This dissertation focuses on the capital cost component to derive a cost per mile at the planning level, and it is based on the statistical relationship and significance of the response variable of cost per mile to predictor variables of multiple alignment characteristics. Based on the wide range of cost per mile estimates for LRT systems, a primary objective of this research is to develop a planning tool that accounts for this variability in cost estimates.

In 2016, North American transit agencies are expected to open 245 miles of new fixed-guideway transit lines (Figure 2.4), including 89 miles of bus rapid transit, 93 miles of commuter rail, 7 miles of heavy rail, 39 miles of light rail, and 18 miles of streetcars (Freemark, 2016).

Figure 2.4: New or Expanded Transit Lines – 2016

Source: Openings and Construction Starts Planned for 2016, Yonah Freemark, The Transport Politic, January 6, 2016

The average cost per mile expected to be completed in 2016 by mode include:

- $4 million for bus rapid transit
- $30 million for commuter rail
- $778 million for heavy rail (two projects)
- $141 million for light rail
- $46 million for streetcar

2.5 Capital Cost Estimating Process for FTA New Starts

The current State of the Practice for the development of transit project capital cost estimates involves adherence to the FTA processes. The FTA implemented the SCC in 2005 for the purposes of establishing a consistent format for the reporting, estimating, and managing of capital costs for New Starts projects. The cost information gathered from projects across the country has been developed into the Capital Cost Database (CCD). The SCC and CCD cost-estimating tools are very relevant to this dissertation's research focus to improve the reliability and accuracy of developing planning level LRT capital cost estimates. They are briefly described below.

2.5.1 Standard Cost Categories for Capital Projects

Concerning cost estimating methods, the FTA website explains that FTA utilizes the SCC to establish a consistent format for the reporting, estimating, and managing of capital costs for New Starts projects. The cost information gathered from projects across the country was intended to generate a database and a cost estimating resource that would be useful to the FTA and the transit industry (FTA, 2016).

The transit agencies follow the most current reporting instructions from Section 5309 (Capital Investment Grant Program, New Starts, August 2015) to report their project costs for federal funding eligibility (FTA Final Policy Guidance, 2013). The SCC format and workbook are explained in Chapter 3.

The SCC worksheets provide a fundamental project management tool for New Start projects and are an important part of operationalizing the Standard Cost Categories. They are also helpful in creating a simple interface with other FTA and project sponsors' funding, budget and grant programs. Project sponsors are required to submit capital cost information electronically in the SCC Excel format. Additionally, the SCC structure accommodates all project elements within 10 major cost categories (FTA, 2016). This dissertation utilizes the SCC format and cost category structure as the basis for the statistical analyses variable development and these analyses.

2.5.2 Capital Cost Database

The current State of the Practice for the development of conceptual or planning level transit project capital cost estimates includes use of the FTA CCD. The CCD is a Microsoft Access database of as-built costs for 54 federally funded projects in the following modes: Bus Rapid Transit, Commuter Rail, Light Rail, Heavy Rail, and Trolley. The projects' costs are tracked in the FTA SCC, and they have been validated.

The purpose of the CCD is to provide the industry with a cost database to document "as-built" costs for several LRT and HRT projects implemented in the U.S. within the last 30 years. The CCD includes project costs and unit quantities recorded at the SCC level of detail. One can also use the CCD to conduct historical cost analysis based on the project costs recorded in the database. The CCD can be used to develop conceptual, "order-of-magnitude" cost estimates for LRT and HRT projects. It does not prepare a detailed cost estimate. The CCD can be used for preparing conceptual estimates for projects or for better understanding the unique characteristics of a cost estimate by comparing the costs to past experiences (FTA, 2016).

While cost model use is initiated from MS Access, the actual cost model resides in MS Excel. Cost estimates produced are sensitive to the cost data included in the user developed “cost basis”. The users cost basis should be obtained from completed projects with similar location context in the database. The cost basis establishes which unit cost data the cost model will draw on from the database in assessing the cost of proposed projects. The cost model is then exported to Excel where the user specific quantity data is used to develop the conceptual cost estimates. Per “*The Capital Cost Database Quick Guide***”** (FTA, 2016), unit cost values of a number of SCC cost elements are modeled using non-linear cost functions that yield decreasing unit cost estimates as larger quantities are entered and users should be cautious of the cost estimates for these cost elements when entering very small unit quantities. Additionally, some SCC categories should only be considered broadly representative of the “expected” costs for these elements based on the experience of the group of projects selected in the cost basis. As with other conceptual cost estimating methodologies, developing a more accurate assessment of the costs for some elements (e.g., right-of-way) requires greater detailed of the specific characteristics of a an alignment.

The CCD is a very useful tool for developing “order-of-magnitude” cost estimates for LRT and HRT projects. It does require the availability of more detailed data (quantities) than would be required for this dissertations’ model. The focus of this dissertation is on the development a statistical regression model for predicting future LRT cost per mile when very limited detailed quantity data has yet to be developed, while the CCD utilizes Excel to produce the estimates for some of the SCC categories that represent an average cost per length of guideway for these types of project expenses.

The CCD structure and how it is utilized in this research is described in Chapter 3.

2.6 Summary

Much of the previous research has produced sound methodologies and practical uses for cost-estimating during early project phases and will be considered and/or utilized in this research. This dissertation cannot resolve many of the limitations for developing capital cost estimates identified above (defined scope, project uncertainties, accuracy of unit cost data, and unavailability of project detail/data). However, this dissertation strives to develop an improved and simplified tool for use at the planning level for LRT capital cost estimates as a component of the evaluation process.

The literature review supports the need for additional research related to the development of methodologies for producing transit capital cost estimates, specifically estimates used for LRT systems, at the planning stage of major capital transportation project development. The review helps to develop the framework for the research methodology described in the next section. Additionally, for the purpose of providing a complete picture of the transit project evaluation process, the review provides a broad overview of the transportation improvement and development process and examines why the development of accurate planning level capital cost element plays a critical role in the overall evaluation and decision-making process. The review supports the importance of cost in the evaluation of alternatives and the decision-making process. Lastly, the literature has shown that this dissertation research will provide a useful tool to transportation planners for developing reliable LRT capital cost estimates at the corridor level to be used in the early stages of project development.

3. METHODOLOGY AND DATA COLLECTION

The development of the statistically based capital cost estimating methodology for transit corridor alignments requires extensive data collection, assessment and preparation. This chapter provides a summary of the methodologies used to accomplish the research objective tasks, which include: transit mode review and selection, data source collection and preparation, LRT project selections/ sample size, and the statistical analysis methods employed.

3.1 Transit Mode Reviews and Selection

Prior to finalizing transit project evaluation and model development methodologies, the selection of the transit technology (mode) needed to be selected in order to focus the data collection and cost analysis. Based on the review of the transit technologies summarized below, the LRT mode was selected for this dissertation research. A summary of this review follows.

Transit technologies can be categorized into several classifications, each of which has particular characteristics that meet certain requirements. The different transit technologies were designed and developed to serve a variety of mobility needs and settings (ITC, 2003). For example, the Local Bus category is best suited for short distance travel in low-density urban areas. Conversely, Automated Guideway Transit (AGT) is best suited for high-density urban areas or activity centers, like central business districts or airports. For medium distance travel in urban and suburban areas Express Buses, Busways, LRT, and HRT are appropriate candidate technologies for implementation. For long suburban-to-urban core trips, or inter-city travel, Commuter Rail Transit (CRT) or High-Speed Rail (HSR) may be the best-suited technologies.

Urban transportation modes are individual modes of transport in urban areas, including public transportation modes. As defined by the federal government, public transportation is “transportation by a conveyance that provides regular and continuing general or special transportation to the public” (APTA, 2005).

Urban transportation modes include:

- Walking
- Bicycles
- Motorcycles and Scooters
- Automobiles
- Paratransit
- Buses (Local/Express, Bus Rapid Transit, and Trolley Bus)
- Streetcars and Light Rail Transit
- Heavy Rail or Rapid Rail Transit
- Commuter Rail Transit

- Automated Guideway Transit
- Monorail
- Water-borne Modes
- Special modes (Cable Cars, Aerial Tramways, Inclined Plane Railways, etc.)

The primary modes under consideration for this research were the conventional line-haul transit systems that are typically considered for urban metropolitan corridors. These transit modes also have the greatest availability and consistency of cost data. The conventional line-haul transit systems category includes primarily HRT/Rapid Transit Systems (either elevated or underground), LRT, BRT, CRT, and Diesel Multiple Unit (DMU) technologies (which either fall under LRT, DLRT or Commuter Rail). In recent years, short portions of light rail systems, particularly in the US, have been elevated. Therefore, at-grade LRT is separated from fully elevated LRT. The transit technologies are defined below (ITC, 2003):

- **Heavy Rail Transit / Rapid Transit Systems –** HRT, also known as Rapid Transit, is a class of high-capacity, urban transport that uses electrically powered railcars operated in long trains over fixed, railroad tracks on an exclusive right-of-way, either in tunnels or on elevated guideways. Large, heavy rail, single, non-articulated cars may be configured in married pairs, where the two cars may share selected common equipment. Service is confined to the corridor with trains stopping frequently at on-line stations located on the main line. Level platform passenger boarding is used, and power is delivered by a wayside third rail arrangement. Because of construction complexities in dense urban areas, costs can be high, but often comparable to light rail projects.

- **At-Grade Light Rail Transit Systems -** LRT is a rail mode comprised of vehicles with steel wheels operating on steel tracks. It is applicable to an entire range of operations, from traditional street installation to an exclusive right-of-way. Modern at-grade implementations of light rail usually avoid mixed LRT/auto use of the track area, and provide an exclusive, in-street right-of-way for the LRT. Traffic operations and safety are issues that need to be addressed in the engineering of at-grade LRT. LRT vehicles are electrically powered, usually by means of overhead wires. Train operation in rush periods typically consists of up to three connected vehicles. Cars may be articulated, and passenger loading is possible at both high-level and low-level platforms. Low-floor vehicles have emerged in recent years.

- **Grade-Separated LRT Systems -** LRT can be utilized in underground applications such as subways, at-grade, or on elevated structures. However, both subway and elevated configurations are not well suited for systems with overhead traction. The elevated LRT guideway structure enables a higher average speed system with a more reliable schedule because there are no conflicts with street traffic or pedestrians. However, the elevated structures often have fewer right-of-way and street impacts than at-grade systems; they also create more visual and aesthetic concerns in comparison with slimmer Rapid Transit guideways. As a general rule, LRT should not be considered in an exclusive elevated configuration, as Rapid Transit Systems offer much better performance.

- **Commuter Rail Systems –** CRT technology resembles intercity rail services, but in urban transit applications. It typically connects suburban areas to a limited number of stations in a downtown area. Commuter rail systems can share the same tracks with regular intercity passenger and freight rail systems. As a result, Commuter Rail must meet the standards and codes of the Association of American Railroads (AAR) and the Federal Railroad Administration (FRA). Commuter Rail systems can be locomotive-hauled or self-propelled vehicles, either diesel or electric. A push/pull commuter train is a locomotive-hauled train capable of operating from either end to facilitate end-off-line turnbacks. For trains that do not have a locomotive at each end, the end passenger car is equipped with an operator's cab that can remotely control the locomotive at the other end of the train. From the operational point of view, trip durations are long in comparison with other transit modes and stations are widely spaced.

- **DMU Rail Systems –** DMU trains, equipped in diesel propulsion, are the most popular types of railway vehicles on non-electrified lines. DMU is a compromise between buses and heavy coaches used on luxury trains. Typically, DMUs are not as comfortable and do not ride as well as locomotive-hauled coaches; however, they are cheaper to maintain and offer more operational flexibility. DMU trains can be joined together with greater ease and in less time than locomotive-hauled trains; additionally, they can also be reversed easier. The majority of DMUs have a top speed of 75 mph, with newer models offering 90 mph. Because of the DMU's typically lower power-to-weight ratio, its acceleration and speed decreases especially on hilly terrain.

- **Bus Rapid Transit Systems -** BRT systems combine the quality of rail transit and the flexibility of buses. They can operate on exclusive transit ways, HOV lanes, expressways, or ordinary streets. A BRT system combines intelligent transportation systems technology, priority for transit, cleaner and quieter vehicles, rapid and convenient fare collection, and integration with land use policy. BRT systems give priority to transit vehicles, since on average, they carry many more people than other road vehicles. One form of priority is to run service on exclusive rights-of-way such as busways and exclusive lanes on expressways. In addition, these techniques can greatly reduce in-vehicle travel time. Another form of priority is to designate bus lanes on arterial streets. Providing traffic signal priority to transit vehicles can also speed operation on streets. Additionally, reducing the number of stops, providing limited-stop service, or relocating stops to areas where there is less congestion can also expedite service, although potentially with the disadvantage of increasing walk time.

The primary source for planning-level and final-built transit cost data for this research is the FTA historical and as-built cost data from FTA-funded and constructed New Start Projects. The CCD includes as-built costs for 54 projects and five transit technology modes. Fifty percent of these projects are LRT. The database breakdown by modes is presented in Table 3.1.

Table 3.1: Project Breakdown by Mode – FTA's Capital Cost Database

Mode	Project Number
Bus Rapid Transit	3
Commuter Rail	5
Heavy Rail Transit	18
Light Rail Transit	27
Streetcar / Trolley	1
Total	54

Many major metropolitan areas that are currently planning for the implementation of major transit investments begin the process considering both BRT and LRT. However, based on the current popularity of the LRT mode and the greater availability of cost data, this research investigates the most frequently encountered LRT mode for the development of a capital cost-estimating methodology at the planning level.

3.2 Data Sources and Collection

This section describes the data sources, collection and format for the model development. The data sources and collection for the assessment and comparison to the LRT systems' planning and conceptual engineering cost estimates developed in the project development phases of AA, EIS, PE, FD and/or at the time of the FFGA are also presented. Finally, the as-built LRT project costs and information regarding the FTA CCD are provided.

3.2.1 Data Format - Standard Cost Categories

The FTA's capital costing format for reporting transit cost is the SCC, which establishes a consistent format for the reporting, estimating, and managing of capital costs for New Starts projects.

Project sponsors are required to submit capital cost information electronically in the SCC Excel format. Cost assumptions from the most current FTA reporting instructions (FTA, 2015) are summarized below:

> “Cost Estimating Assumptions: A project’s capital cost estimate includes costs for planning, design and construction. It includes labor and materials for construction of the improvement – such as guideways, stations, support facilities, sitework, special conditions and systems – as well as costs for vehicle design and procurement, environmental mitigation, right-of-way acquisition, relocation of existing households and businesses, planning, facility design, construction management, project administration, finance charges, and contingencies. New Starts project sponsors must use the most recent SCC worksheets issued by FTA for reporting the capital costs and schedules of their proposed projects. New Starts project sponsors should report costs in reporting 2015 constant dollars…FTA expects that cost estimates for the project be up-to-date, be based on unit costs that apply to expected conditions during construction, and specifically identify remaining uncertainties in those unit costs”.

A data request was made to the FTA in order to receive the most relevant and up-to-date LRT cost data available for distribution to be used in this research. The initial request yielded LRT cost data in SCC format from 17 LRT systems. The FTA has recently updated the CCD (May 2016) with additional projects so a supplemental request was submitted, and data for 10 additional LRT systems were provided.

The SCC structure, which includes all project cost elements within 10 major cost categories, is presented in Table 3.2. A more detailed list that includes the SCC definitions, as well as a sample “Main Worksheet” for a “Build Alternative,” is included in Appendix 3-A.

3.2.2 Capital Cost Database

The FTA CCD is a database of as-built costs for completed FTA-funded major transit projects. Project costs are in SCC format and have been confirmed by project sponsors. The CCD includes as-built costs for the 27 LRT projects completed over the last 20 years.

The purpose of the CCD is to develop conceptual, “order-of-magnitude” cost estimates for potential projects in the modes listed above. The CCD is initiated from Microsoft Access, but the cost model is in Microsoft Excel. The LRT as-built cost data received from the FTA for this dissertation are included in the current CCD. The final as-built cost for LRT projects were collected from the CCD.

Table 3.2: Standard Cost Categories for New Starts Projects

Standard Cost Categories for New Starts Projects

(Rev.17, June, 2015)

10 GUIDEWAY & TRACK ELEMENTS (route miles)

10.01	Guideway: At-Grade Exclusive Right-Of-Way
10.02	Guideway: At-Grade Semi-Exclusive (Allows Cross-Traffic)
10.03	Guideway: At-Grade in Mixed Traffic
10.04	Guideway: Aerial Structure
10.05	Guideway: Built-Up Fill
10.06	Guideway: Underground Cut and Cover
10.07	Guideway: Underground Tunnel
10.08	Guideway: Retained Cut or Fill
10.09	Track: Direct Fixation
10.10	Track: Embedded
10.11	Track: Ballasted
10.12	Track: Special (Switches, Turnouts)
10.13	Track: Vibration and Noise Dampening

20 STATIONS, STOPS, TERMINALS, INTERMODAL (number)

20.01	At-Grade Station, Stop, Shelter, Mall, Terminal, Platform
20.02	Aerial Station, Stop, Shelter, Mall, Terminal, Platform
20.03	Underground Station, Stop, Shelter, Mall, Terminal, Platform
20.04	Other Stations, Landings, Terminals: Intermodal, Ferry, Trolley, etc.
20.05	Joint Development
20.06	Automobile Parking Multi-Story Structure
20.07	Elevators, Escalators

30 SUPPORT FACILITIES: YARDS, SHOPS, ADMIN. BLDGS

30.01	Administration Building: Office, Sales, Storage, Revenue Counting
30.02	Light Maintenance Facility
30.03	Heavy Maintenance Facility
30.04	Storage or Maintenance of Way Building
30.05	Yard and Yard Track

40 SITEWORK & SPECIAL CONDITIONS

40.01	Demolition, Clearing, Earthwork
40.02	Site Utilities, Utility Relocation
40.03	Hazardous Material, Contaminated Soil Removal/Mitigation, Ground Water Treatments
40.04	Environmental Mitigation, e.g. Wetlands, Historic/Archeological, Parks
40.05	Site Structures Including Retaining Walls, Sound Walls
40.06	Pedestrian / Bike Access and Accommodation, Landscaping
40.07	Automobile, Bus, Van Access Ways Including Roads, Parking Lots
40.08	Temporary Facilities and Other Indirect Costs During Construction

Table 3.2 (continued)

50 SYSTEMS		
	50.01	Train Control and Signals
	50.02	Traffic Signals and Crossing Protection
	50.03	Traction Power Supply: Substations
	50.04	Traction Power Distribution: Catenary and Third Rail
	50.05	Communications
	50.06	Fare Collection System and Equipment
	50.07	Central Control
60 ROW, LAND, EXISTING IMPROVEMENTS		
	60.01	Purchase or Lease of Real Estate
	60.02	Relocation of Existing Households and Businesses
70 VEHICLES (number)		
	70.01	Light Rail
	70.02	Heavy Rail
	70.03	Commuter Rail
	70.04	Bus
	70.05	Other
	70.06	Non-Revenue Vehicles
	70.07	Spare Parts
80 PROFESSIONAL SERVICES (applies to Cats. 10-50)		
	80.01	Project Development
	80.02	Engineering
	80.03	Project Management for Design and Construction
	80.04	Construction Administration & Management
	80.05	Professional Liability and other Non-Construction Insurance
	80.06	Legal; Permits; Review Fees by Other Agencies, Cities, etc.
	80.07	Surveys, Testing, Investigation, Inspection
	80.08	Start Up
90 UNALLOCATED CONTINGENCY		
100 FINANCE CHARGES		

Source: SCC_Workbook_Rev_17_NEW_START_(FINAL) – Microsoft Excel

3.3 Statistical Analysis Methods

The statistical tool selected for use in this research was the JMP software application of Statistical Analysis Systems (SAS), (Lehman, 2005), because of the ease of use and ability to combine statistical analyses with graphical interface to display and analyze the data.

The models were developed through a comprehensive analysis of the data and statistical techniques including: analysis of the SCC format project cost data, distribution analysis, bivariate relationship analysis, correlation analysis, and regression analysis. These techniques are briefly described below and elaborated on in Chapter 4.

- **SCC project cost data analysis –** The SCC data included 10 cost categories (see Table 3.2) which consisted of 60 sub-categories. These sub-categories included over 150 sub-elements of cost data for each of the 27 LRT projects. The data was reviewed to determine which of the over 150 variables were the best predictors of cost per mile, identify missing data, and determine what variables could potentially be combined and/or eliminated from the final data set.
- **Distribution analysis –** The distribution of the data was examined as part of the data screening process to ensure that data was entered correctly before conducting more advanced statistical techniques. The analysis was also performed to investigate the distribution of the cost data. Multiple linear regression analysis requires that the data be drawn from a normally distributed population. Normality testing was conducted and the analysis indicated some positive and negative skewedness (abnormal distributions) in some of the cost data that later required log transformations (described in Chapter 4).
- **Bivariate analysis –** Procedures were performed to test for the significance of the relationship between variable pairs. The bivariate relationships examined the relationship between pairs of the independent (predictor) variables to determine their independence from each other. Additionally, the bivariate relationship was investigated for the dependent (response) variable, cost per mile, with all other independent cost variables. Scatter plots were developed to investigate linearity between the two variables. This was done because multiple linear regression analysis requires that the relationship between the response variable and the predictor variables be linear. The scatter plots were also useful to check for potential outlier effects in the data.
- **Correlation analysis –** When using a least squares regression procedure, the total number of predictor variables needs to be less than the number of observations. Correlation analysis of the variables was performed to review the correlation of the predictor variables so that highly correlated variables were not used in the model development testing. Multicollinearity occurs when the predictor (independent) variables are not independent from each other, and those variables should not be used in linear regression. The Pearson Correlations and Spearman Correlations were considered.
- **Regression analysis -** Multiple linear regression techniques can be used to assess whether one continuous response variable (cost per mile) can be estimated from a group of predictor variables. A selection model procedure was utilized to reduce the number of predictor variables that was needed to account for the variance of the total group of predictors. The step-wise selection procedure was primarily used in developing the model. The step-wise selection procedure is a variation of both the forward selection and backwards elimination methods. It includes analysis at each step to determine the contribution of the predictor variable entered into the equation. This method provided insight into the contributions of statistical significance of the previous variables. Testing involved the adding, deleting and retaining of variables based on their statistical significance.

3.4 Project Selections and Sample Size

The LRT mode was selected for development of the cost model based on the review of transit technologies summarized above, the current popularity of the LRT mode and the greater availability of cost data available from the data sources described above. The final project list includes a sample size of 27 LRT projects that were entered into revenue service between 1986 and 2012. Table 3.3 presents the projects.

Table 3.3: LRT Projects Included in Sample Size

Project Name	Location	Alignment Length (Miles)
1 - Charlotte - South Corridor LRT	Charlotte, NC	8.7
2 - Denver – SE Corridor Project T-Rex	Denver, CO	19.0
3 - Denver - Southwest Corridor	Denver, CO	7.0
4 - Los Angeles - East Side Extension	Los Angeles, CA	4.3
5 - Los Angeles - Long Beach Blue Line	Los Angeles, CA	22.6
6 - Minneapolis - Hiawatha Corridor	Minneapolis, MN	11.6
7 - New Jersey - Hudson-Bergen MOS-2	Newark, NJ	6.0
8 - New Jersey - Newark Rail Link	Newark, NJ	8.8
9 - New Jersey - Southern NJ LRT System	Trenton, NJ	28.0
10 - Norfolk - Light Rail Transit	Norfolk, VA	6.8
11 - Phoenix - Central Phoenix/East Valley	Phoenix, AZ	39.4
12 - Pittsburgh - Light Rail Stage I	Pittsburgh, PA	15.6
13 - Pittsburgh - Light Rail Stage II	Pittsburgh, PA	5.4
14 - Pittsburgh - North Shore Connector	Pittsburgh, PA	1.2
15 - Portland - Interstate MAX	Portland, OR	5.8
16 - Portland - MAX Segment I	Portland, OR	19.6
17 - Portland – So. Corridor/Portland Mall	Portland, OR	7.6
18 - Portland - Westside/Hillsboro MAX	Portland, OR	17.7
19 - Sacramento - Folsom Corridor	Sacramento, CA	12.9
20 - Sacramento - South Corridor	Sacramento, CA	6.3
21 - Sacramento - Stage I	Sacramento, CA	21.2
22 - Salt Lake City - Mid Jordan LRT	Salt Lake City, UT	21.0
23 - Salt Lake City - North South Corridor	Salt Lake City, UT	15.1
24 - San Diego - Mission Valley East	San Diego, CA	5.5
25 - Santa Clara VTA - North Corridor	San Jose, CA	15.6
26 - Santa Clara VTA - Tasman West	San Jose, CA	17.5
27 - St. Louis - St. Clair County Extension	St. Louis, MO	7.4

4. DATA ANALYSIS AND PLANNING LEVEL LRT CAPITAL COST MODEL DEVELOPMENT

The development of the statistical based planning level LRT capital cost estimating model included extensive data collection, preparation, and analysis. This chapter discusses the LRT systems' alignment characteristics cost data analysis and model variable development from the source cost data, cost data adjustments from the LRT system implementation year dollars to 2015 dollars, the model development and testing activities, and a summary of the final multiple linear regression model results.

4.1 Data Analysis and Variable Development

As noted above, the FTA database provided cost data for 27 federally funded, constructed LRT systems in the United States. The data included both cost (LRT revenue year dollars – the year the system entered into revenue service) and quantities in an Excel file for 8 of the 10 SCC categories. As described previously (see Table 3.2), the SCC is the FTA capital costing format structure for reporting transit cost data. Additionally, each of the 8 SCC categories of separate costs, which included 56 sub-categories, were included in the data files for each LRT system. All SCC data was examined first to determine their potential for inclusion in the model as predictor variables. The SCC structure (categories and sub-categories) included 56 variables that were the best potential predictors of project cost. An important step in the variable development was the review of the source data to gain an understanding of which of the 56 variables could be used to best predict cost per mile for the LRT systems. Four basic criteria were initially used to determine which of the variables could be used for potential model development and testing. The criteria are as follows: 1) Did the variable have sufficient data for each of the LRT systems? 2) Could the variable data potentially be quantified during the planning phase of project development? 3) Is the variable data mutually exclusive? 4) Could the data from multiple variables (sub-categories) be aggregated into a single variable?

Utilizing these criteria, the number of potential data variables (predictors) for model development was reduced from 56 to 44. The 44 variables are presented and described in Table 4.1.These 44 variables were analyzed further for use in potential model development as described in Section 4.3 below.

Table 4.1: Potential Data Variables for Model Development

Variable Number	Variable Name	Systems with Data	Unit	SCC Category	SCC Definitions
1	Alignment_Length	27	Miles		
2	Guideway_01_Q	22	LF Guideway	10.01	Guideway: At-grade exclusive right-of-way
3	Guideway_01_$	22	2015 Dollars	10.01	Guideway: At-grade exclusive right-of-way
4	Guideway_02_Q	8	LF Guideway	10.02	Guideway: At-grade semi- exclusive (allows cross-traffic)
5	Guideway_02_$	8	2015 Dollars	10.02	Guideway: At-grade semi- exclusive (allows cross-traffic)
6	Guideway_03_Q	10	LF Guideway	10.03	Guideway: At-grade in mixed traffic
7	Guideway_03_$	10	2015 Dollars	10.03	Guideway: At-grade in mixed traffic
8	Guideway_04_Q	17	LF Guideway	10.04	Guideway: Aerial
9	Guideway_04_$	18	2015 Dollars	10.04	Guideway: Aerial
10	Guideway_05_Q	7	LF Guideway	10.05	Guideway: Built-up fill
11	Guideway_05_$	8	2015 Dollars	10.05	Guideway: Built-up fill
12	Guideway_06_Q	5	LF Guideway	10.06	Guideway: Underground cut &
13	Guideway_06_$	5	2015 Dollars	10.06	Guideway: Underground cut &
14	Guideway_07_Q	4	LF Guideway	10.07	Guideway: Underground tunnel
15	Guideway_07_$	5	2015 Dollars	10.07	Guideway: Underground tunnel
16	Guideway_08_Q	11	LF Guideway	10.08	Guideway: Retained cut or fill
17	Guideway_08_$	11	2015 Dollars	10.08	Guideway: Retained cut or fill
18	Guideway_09_Q	16	Track Feet	10.09	Track: Direct fixation
19	Guideway_09_$	16	2015 Dollars	10.09	Track: Direct fixation
20	Guideway_10_Q	13	Track Feet	10.10	Track: Embedded
21	Guideway_10_$	13	2015 Dollars	10.10	Track: Embedded
22	Guideway_11_Q	20	Track Feet	10.11	Track: Ballasted
23	Guideway_11_$	20	2015 Dollars	10.11	Track: Ballasted
24	Guideway_12_Q	16	Track Feet	10.12	Track: Special (switches,turnouts)

Table 4.1 (continued)

Variable Number	Variable Name	Systems with Data	Unit	SCC Category	SCC Definitions
25	Guideway_12_$	17	2015 Dollars	10.12	Track: Special (switches, turnouts)
26	Station_01_Q	23	Stations	20.01	At-grade station, stop, shelter, mall, terminal, platform
27	Station_01_$	23	2015 Dollars	20.01	At-grade station, stop, shelter, mall, terminal, platform
28	Station_02_Q	4	Stations	20.02	Aerial station, stop, shelter, mall,
29	Station_02_$	4	2015 Dollars	20.02	Aerial station, stop, shelter, mall,
30	Station_03_Q	7	Stations	20.03	Underground station, stop, shelter, mall, terminal, platform
31	Station_03_$	7	2015 Dollars	20.03	Underground station, stop, shelter, mall, terminal, platform
32	Station_06_Q	5	Spaces	20.06	Automobile parking multi-story structure
33	Station_06_$	6	2015 Dollars	20.06	Automobile parking multi-story structure
34	Support_Fac_$	21	2015 Dollars	30.01-.05	30 SUPPORT FACILITIES: YARDS, SHOPS,
35	Sitework_$	23	2015 Dollars	40.01, .05, .08	Demolition, Clearing, Earthwork; Site structures including retaining walls, sound walls; Temporary Facilities and other indirect costs during construction
36	Utilities_$	25	2015 Dollars	40.02	Site Utilities, Utility Relocation

Table 4.1 (continued)

Variable Number	Variable Name	Systems with Data	Unit	SCC Category	SCC Definitions
37	Environmental_$	18	2015 Dollars	40.03 - .04	Haz. mat'l, contam'd soilremoval/mitigation, ground water treatments; Environmental mitigation, e.g. wetlands, historic/archeologic, parks
38	Accessways_$	14	2015 Dollars	40.06 - .07	
39	Systems_$	25	2015 Dollars		
40	ROW_Q	27	LF Guideway	60.01	Purchase or lease of real estate
41	ROW_$	27	2015 Dollars	60.01	Purchase or lease of real estate
42	Vehicles_Q	25	Vehicles	70.01	Light Rail
43	Vehicles_$	26	2015 Dollars	70.01	Light Rail
44	Prof_Serv_$	27	2015 Dollars	80.01 - .08	Project Development

Notes: Variable names with "_Q" are quantity variables; Variable names with "_$" are Revenue opening year $ converted to 2015 $. LF = Linear Feet.

4.2 Cost Data Adjustments

The cost data for the 27 LRT projects used in the model development commenced revenue service between 1986 and 2012. Due to inflation and other factors, cost will vary over time. Since the objective of the final model is to predict future LRT system's cost per mile, the cost from the opening revenue year (final built construction cost in year of expenditure dollars) should be escalated to a consistent future year for all projects.

Prior to performing the statistical data analysis for the variables, the opening revenue year cost was adjusted to 2015 dollars for each LRT system. The cost adjustments were made with the RS Means Construction Cost Indexes based on an historical index. A table that lists both the RS Means historical cost index based on January 1, 1993 as 100 and the computed value of an index based on January 1, 2016 costs is in Appendix 4-A. Table 4.2 includes the LRT projects revenue year cost, adjusted 2015 cost, and time adjustment indices.

4.3 Model Development and Testing

This section describes the statistical data analyses performed for the model development and testing.

4.3.1 Distribution Analysis

The distribution technique was used for data descriptive analyses purposes. It was useful to investigate the distribution of the selected 44 predictor variables and results to ensure that the data looked reasonable (no apparent errors in data entry, impossible values such as negative cost values, and shape of data) before performing more complex statistical analyses. The analysis was also performed to investigate the shape of the cost data; multiple linear regression analysis requires that the data be drawn from a normally distributed population.

The distribution technique analyzed the 44 quantitative (numeric) variables presented in Table 4.1 for all 27 LRT projects. The number of LRT systems that each of the 44 variables has data (sample size for each variable) is included in the table. Appendix 4-B provides the detailed data for each of the variables considered. The analysis provided the mean, standard deviation, standard error of the mean, as well as the upper and lower confidence levels (95% mean). A sample distribution for the at-grade station variable (quantity) is shown in Figure 4.1. The distribution graphs for all variables are included in Appendix 4-C.

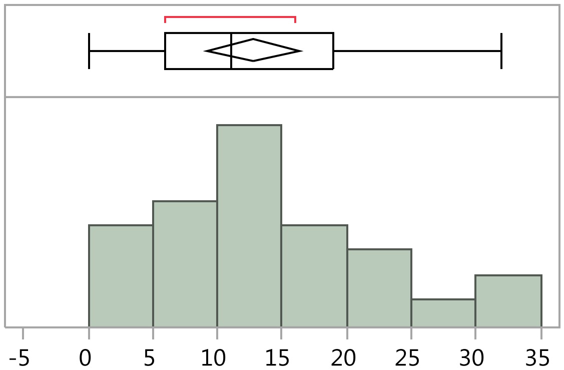
Table 4.2: Cost Data Adjustments							
Project #*	Total Project Cost (Revenue Year)	Cost per Mile (Revenue Year)	Total Project Cost (2015 Dollars)	Cost per Mile (2105)	Open Revenue Year	RSMeans Const Cost Index	Time Adjustment Indices (2015/cost yr)
1	$462,748,292	$48,202,947.08	$563,274,485.30	$58,674,425.55	2007	169.4	1.22
2	$878,959,094	$46,261,004.95	$1,118,773,859.15	$58,882,834.69	2006	162.0	1.27
3	$177,100,001	$20,835,294.24	$302,051,449.18	$35,535,464.61	2000	120.9	1.71
4	$876,079,616	$141,303,163.87	$1,003,040,626.43	$161,780,746.20	2009	180.1	1.14
5	$877,269,855	$38,817,250.22	$1,918,271,941.69	$84,879,289.46	1990	94.3	2.19
6	$672,477,878	$57,972,230.86	$964,961,297.45	$83,186,318.75	2004	143.7	1.43
7	$1,006,165,000	$167,694,166.67	$1,404,679,911.98	$234,113,318.66	2003-06	147.7	1.40
8	$206,212,000	$23,433,181.82	$262,474,780.25	$29,826,679.57	2006	162.0	1.27
9	$698,599,350	$24,949,976.79	$1,091,296,863.41	$38,974,887.98	2003	132.0	1.56
10	$307,894,159	$41,607,318.78	$332,049,035.49	$44,871,491.28	2011	191.2	1.08
11	$1,264,298,130	$62,280,696.06	$1,445,112,385.84	$71,187,802.26	2008	180.4	1.14
12	$555,605,245	$35,615,720.83	$1,306,337,531.57	$83,739,585.36	1987	87.7	2.35
13	$386,157,500	$71,510,648.15	$554,110,483.65	$102,613,052.53	2004	143.7	1.43
14	$502,588,000	$418,823,333.33	$532,546,996.92	$443,789,164.10	2012	194.6	1.06
15	$343,236,000	$59,178,620.69	$492,520,968.68	$84,917,408.39	2004	143.7	1.43
16	$321,313,000	$16,393,520.41	$786,873,403.80	$40,146,602.23	1986	84.2	2.45
17	$551,689,839	$67,279,248.66	$631,640,448.65	$77,029,323.01	2009	180.1	1.14
18	$969,182,332	$54,756,063.95	$1,736,276,254.20	$98,094,703.63	1998	115.1	1.79
19	$268,285,714	$20,797,342.17	$364,911,043.71	$28,287,677.81	2005	151.6	1.36
20	$223,821,859	$35,527,279.21	$349,636,873.68	$55,497,916.46	2003	132.0	1.56
21	$163,636,863	$7,718,719.95	$384,742,544.48	$18,148,233.23	1987	87.7	2.35

Table 4.2 (continued)

Project #*	Total Project Cost (Revenue Year)	Cost per Mile (Revenue Year)	Total Project Cost (2015 Dollars)	Cost per Mile (2105)	Open Revenue Year	RSMeans Const Cost Index	Time Adjustment Indices (2015/cost yr)
22	$480,532,969	$44,493,793.43	$518,231,685.19	$47,984,415.30	2011	191.2	1.08
23	$294,944,466	$19,532,746.09	$517,156,027.97	$34,248,743.57	1999	117.6	1.75
24	$504,014,126	$91,638,932.00	$685,539,002.51	$124,643,455.00	2005	151.6	1.36
25	$339,325,126	$21,751,610.64	$797,820,307.65	$51,142,327.41	1987	87.7	2.35
26	$359,861,719	$47,981,562.53	$593,153,368.97	$79,087,115.86	2001	125.1	1.65
27	$350,602,680	$47,378,740.54	$577,891,867.43	$78,093,495.60	2001	125.1	1.65
RS Means Construction Cost Index for Escalation Year to 2015 Dollars					2015	206.2	1.00

** Project # corresponds to the project name in Table 3.3*

Figure 4.1: Sample Data Distribution – At-Grade Stations



Frequency of At-Grade Stations (Staion_01_Q)

The mean for the quantity of at-grade stations (12.7) falls between the variable range and appears to be correct (Table 4.3). In the Quantile table (Table 4.4), the Minimum represents the lowest value (0) in the data table, which indicates no problems with this variable. The Maximum does not exceed 32, so this confirms there are no errors in the at-grade station variable.

Table 4.3: Sample Distribution Summary Statistics

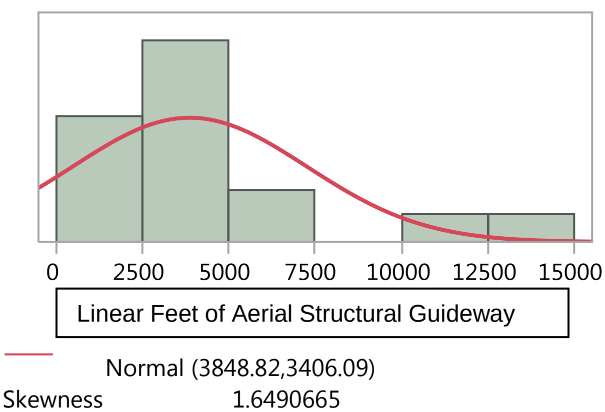
Category	Value
Mean	12.704
Standard Deviation	8.973
Standard Error - Mean	1.725
Upper 95% Mean	16.250
Lower 95% Mean	9.157
Number	27

Table 4.4: Sample Distribution Quantiles

Percentage	Category	Value
100.0%	maximum	32
99.5%		32
97.5%		32
90.0%		28.4
75.0%	quartile	19
50.0%	median	11
25.0%	quartile	6
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

The histograms show that the data from this variable form an approximately normal distribution. Each of the variables was investigated for normality. Normality testing was conducted as part of the data analysis because substantial non-normal data can lead to inaccurate conclusions in inferential statistical analysis, as well as a biasing effect on correlation coefficients. A normal distribution should have a bell shaped (symmetrical) distribution of values. The normality analysis indicated some positive and negative skewness (abnormal distributions) in some of the cost data that later required log transformations. A distribution is skewed if the tail on one side of the distribution is longer than the tail on the other side. A sample distribution with fitted normal and smooth curves on the histogram for the guideway on aerial structure variable (quantity) is shown in Figure 4.2. Both visual and statistical tests for normality were performed. The distribution and fitted normal curve displays a positive skewness since the longer tail of the distribution points in the direction of the higher values, which indicates a departure from normality. A skewness value was calculated and was positive (1.65) confirming that the distribution is positively skewed. Additionally, the Shapiro-Wilk (W) test was used for testing the fit of a normal distribution to the sample. The W statistic was 0.808 with a *p* value 0.003. The very small *p* value provides evidence that one should reject the null hypothesis that the sample data are normally distributed. The distribution with fitted normal and smooth curves for all variables is included in Appendix 4-D.

Figure 4.2: Sample Fitted Normal Curve – Guideway on Aerial Structure



4.3.2 Bivariate and Correlation Analysis

Bivariate analysis involves investigating the relationship between two variables in the data set. Statistical tests can be used to test the significance of the relationship between variables. The first set of statistical tests conducted was to investigate the relationship between all pairs of predictor variables. The correlation analyses of the predictor variables were performed to identify the highly correlated variables for possible elimination in the model. Highly correlated predictor variables should not be used in the model at the same time. Removing highly correlated values can minimize the problems that may occur with multicollinearity, which can exist when two or more of the predictors are highly correlated in a model. This can reduce the degree of confidence in the regression model because effect attributed to each predictor variable will be difficult to separate.

The Pearson or Spearman correlations can be used to test for correlations when the predictor variable is continuous. The Pearson correlation is preferable when the relationship between the two variables is linear, the predictor and response are continuous variables, and either on the interval or ratio level of measurement. The Spearman correlation coefficient is best when the predictor and response variables are on the ordinal level of measurement. The Spearman coefficient is a distribution-free test and can also be used when both variables are continuous but is non-normal in distribution, such as the case with skewed data. The primary correlation analysis used the Pearson correlation, but was cross-checked with the Spearman correlation on a small sample of variables with markedly non-normal data distribution. The Spearman correlations were very consistent and supported the findings of the Pearson correlations regarding both strong and weak pairwise correlations of predictor variables.

The Pearson correlation coefficients were calculated for each pair of predictor variables. The correlation coefficients range in size between -1.00 to +1.00. A coefficient of +/- 1.00 indicates a perfect positive or negative linear relationship and 0.00 indicates no relationship between two variables. A few examples of predictor variable pairs with strong correlation are presented in Table 4.5. For instance, the data shows that the length and value of Guideway_06 are correlated, as it was anticipated since the cost is a function of the length. Similarly, the quantity and cost of Station_02 and the alignment length and Guideway_01 length turned out to be related, which confirmed expectations. These tests identified these pairs of variables that should not be used simultaneously in the model.

Table 4.5: Sample of Variable Pairs with Strong Correlation

Variable Pair	Pearson Correlation Coefficient	Significance Probability (*p*-value)
Guideway_06_$ x Guideway_06_Q	0.961	<.0001
Station_02_Q x Station_02_$	0.807	<.0001
Guideway_01_Q x Alignment_Length	0.702	<.0001

Figure 4.3 shows the scatterplot matrix with a 95% density ellipse overlaid on the pairwise correlation of the Guideway_06 (underground cut & cover) quantity and cost predictor variables. The Pearson correlation coefficient between the two variables is 0.96 which indicates a very strong correlation. If the *p*-value obtained from a test of the null hypothesis has the correlation of zero, it means that there is no relationship (or predictive value) between the variables. The calculated p-

value is less than 0.0001, which means there is less than a 1 in 10,000 chance of obtaining a correlation of 0.96 or larger if the population correlation is zero. Therefore, one can reject the null hypothesis, and assume that Guideway_06_Q and Guideway_06_$ are correlated. An observed correlation is typically considered to be statistically significant if the *p* value is less than 0.05. Based on this strong correlation and the difficulty of quantifying the underground cut & cover guideway cost at the planning level, the Guideway_06_$ will be eliminated from use in the model development. A strong correlation also existed between several other guideway quantity variables and its associated cost variables, so many of the guideway cost variables were eliminated because of the inability to estimate the cost at the planning level, and the guideway quantity variables were used.

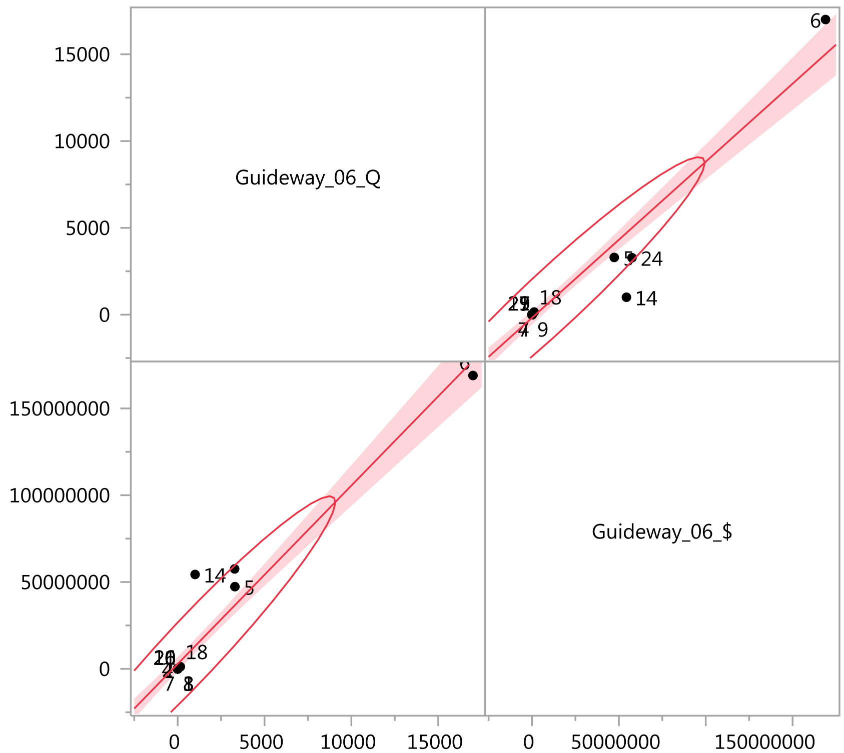
The other relationship example shown in Table 4.5 compares the aerial station variables for quantity (Station_02_Q) and cost (Station_02_$). Based on the high correlation between the two variables, and similarity as discussed above regarding guideway cost estimates during planning phases, the aerial station cost variable was eliminated.

Finally, the third example examined the relationship of the guideway at grade in exclusive right-of way (Guideway_01_Q) and the total route alignment length converted to linear feet (Alignment_Length). As expected, the two variables have a strong correlation. Based on this strong correlation, as well as the fact that the guideway at-grade in exclusive right-of-way (Guideway_01_Q) is a sub-component of the overall total route alignment length (Alignment_Length), only one or the other should be used along with other predictor variables for model development testing. Various model tests were conducted with both variables to determine the best-fit model, but both variables were not used simultaneously in the same model.

A few examples of predictor variable pairs with weak correlation are also presented here. These variable pairs have Pearson correlation coefficients between 0.0 and 0.3 indicating weak correlation, and could be used in conjunction as predictor variables for model development tests. Figure 4.4 shows the scatterplot matrix with a 95% density ellipse overlaid on the pairwise correlation of the quantity underground cut & cover guideway configuration (Guideway_06_Q)) and the underground station variables for quantity (Station_03_Q) predictor variables. The Pearson correlation coefficient between the two variables is 0.23, which indicates a weak correlation and a weak linear relationship.

The other relationship example shown in Table 4.6 compared the at-grade station variables for quantity (Station_01_Q) and at-grade guideway in mixed traffic quantity (Guideway_03_Q). Based on the weak correlation between the two variables, and the ability to estimate their quantities relatively easily during planning phases, these two variables were retained for use in the model development test.

Figure 4.3: Sample Correlation – Guideway: Underground Cut & Cover (quantity by cost)

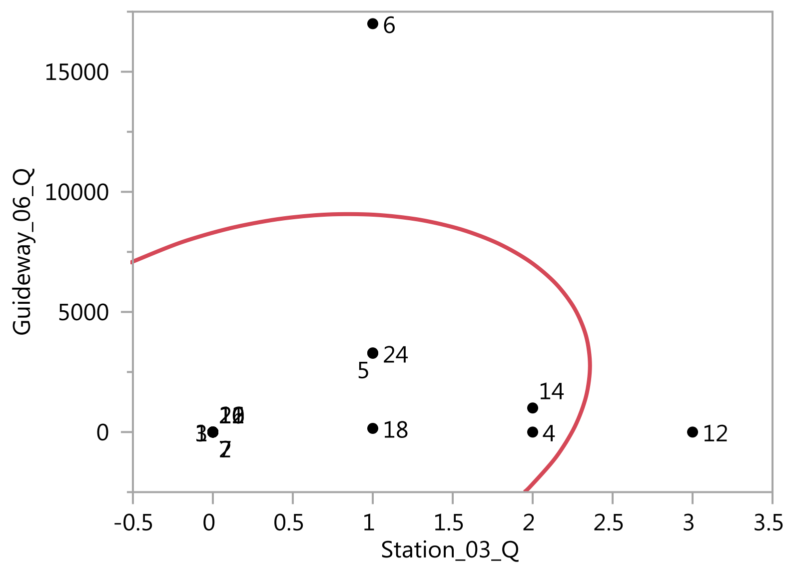


Finally, the third example of a weak correlation examined the relationship between the environmental mitigation cost (Environmental_$) and the right-of-way acquisition quantity (ROW_Q) variables. The two variables have a weak negative correlation and were retained for inclusion in model development.

Table 4.6: Sample of Variable Pairs with Weak Correlation

Variable Pair	Pearson Correlation Coefficient	Significance Probability (*p*-value)
Station_03_Q x Guideway_06_Q	0.227	0.2556
Station_01_Q x Guideway_03_Q	0.206	0.3024
Environmental_$ x ROW_Q	-0.088	0.6692

Figure 4.4: Sample Correlation – Guideway: Underground Cut & Cover by Underground Station



An additional bivariate analysis was conducted to examine the relationship between LRT cost per mile (response) and its potential predictors. Scatterplots were developed to investigate linearity and correlation between the two variables. This scatterplots were created because multiple linear regression analysis requires a linear relationship between the response variable and the predictor variables. The multiple linear regression technique seemed to be the best form to use based on data exploration and because the variables data (numeric ratio scale) has a continuous modeling type. Also, the scatterplots were useful to check for potential outliers in the data.

There is a linear relationship between two variables when the scatterplot follows the form of a straight line. If the scatterplot does not follow the form of a straight line, a nonlinear relationship exists. The linearity assumption is common in the development of regression models for prediction purposes. An illustration of bivariate analysis to examine linearity by generation of scatterplots is shown in Figures 4.5 and 4.6.

These scatterplots use the response variable (cost per mile) plotted on the vertical (y) axis and the predictor variables (underground stations; right-of-way acquisition) on the horizontal (x) axis to explore linearity and correlation. The scatterplot in Figure 4.5 shows the results of Cost per Mile by Underground Stations, and the shape shows a somewhat positive linear relationship. One would expect this, because as the number of expensive underground station increases, so does cost per mile. The scatterplot in Figure 4.6 shows the results of Cost per Mile by Right-of-Way, and the shape shows a somewhat negative relationship. This is somewhat unexpected, because as the amount of right-of-way that would need to be acquired increases, one may expect an increase in cost per mile. One reason this may have a negative relationship with the LRT capital cost data is that the right-of-way required could be relatively inexpensive based on land use type or the right-of-way may have been donated. Additionally, this trend could be due to the available data, since several of the projects required minimal right-of-way. Based on this unexpected relationship and the difficulty in estimating right-of-way acquisition requirements at the planning level, this variable may not be a good predictor variable to use in the model development. This variable's unsuitability was later confirmed during the regression analysis described below.

Figure 4.5: Bivariate Fit of Cost per Mile by Underground Stations

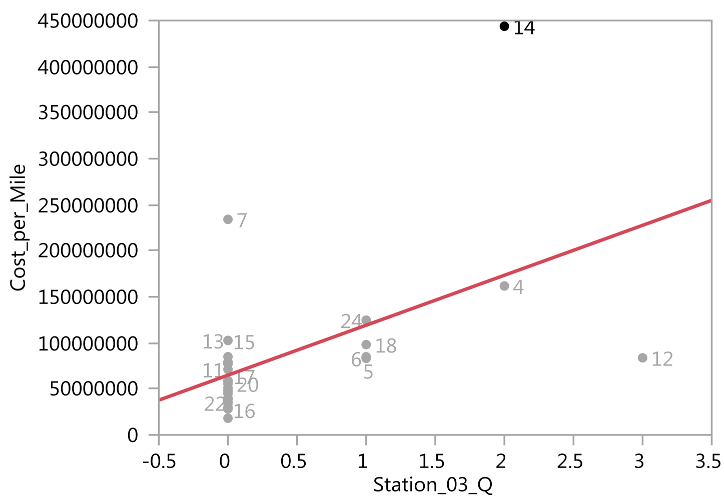
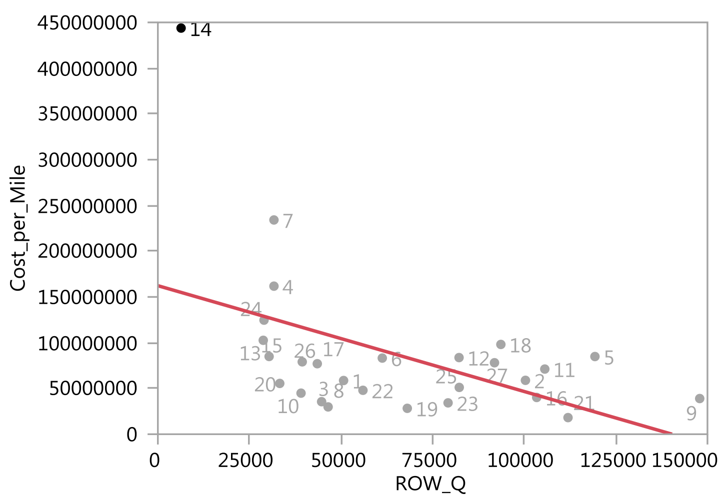


Figure 4.6: Bivariate Fit of Cost per Mile by Right-of-Way



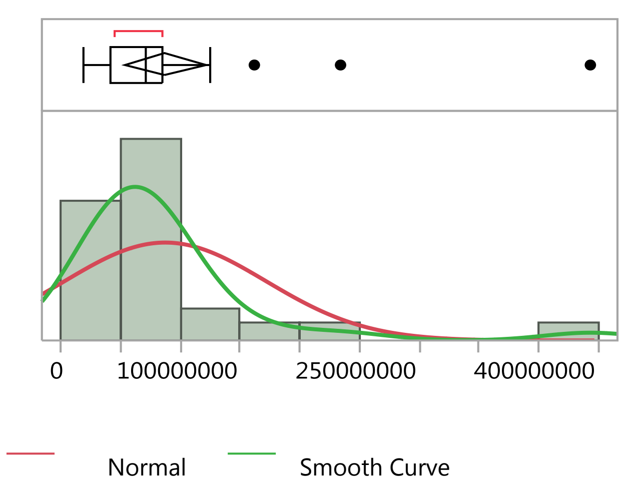
The scatterplots and line fits used in the examination of the linear versus nonlinear relationship for the response variable (cost per mile) by all predictor variables are included in Appendix 4-E**.** This scatterplot analysis was helpful in identifying the strongest possible predictor variables for model testing.

4.3.3 Data Transformations

Normality testing was conducted as part of the data distribution analysis described previously. The analysis indicated some positive and negative skewness (abnormal distributions) in some of the cost data. If a measurement variable does not fit a normal distribution, there may be a need for data transformation.

Using linear regression or another parametric statistical test on data that is not normally distributed could produce misleading results. However, transforming the data may make it fit the model assumptions better. Logarithmic transformations are one method of transforming a highly skewed variable into one that is more approximately normal. Several of the variables were highly skewed. For example, the response variable is highly skewed, (Cost per Mile) in the data set (see Figure 4.7). There is a mix of opinions in the research regarding the need to transform dependent or response variables. The research found that many researchers transform skewed Y distributions before they run the model. The distributional assumptions for linear regression are for the distribution of Y (response variable) given X (predictor variable). Meaning you have to take out the effects of all the Xs before you look at the distribution of Y. In this research, some preliminary sample regression models were tested with both non-transformed and transformed response variables, with the transformed data providing more significant preliminary results. Based on this pre-analysis and the highly skewed variable data, transformations were performed for the regression model development analysis.

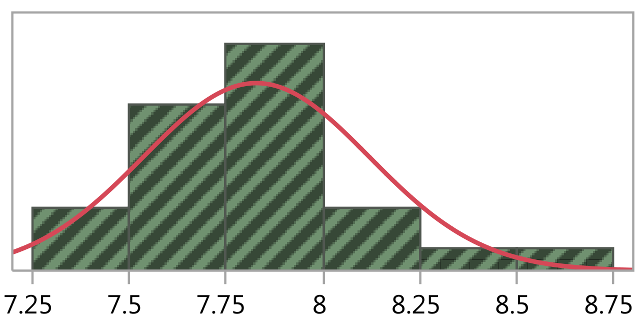
Figure 4.7: Sample of Highly Skewed Variable Data, Cost per mile



The Cost per Mile variable is highly positively skewed to the right, which may cause misleading results if not transformed. Also, several other cost variables did not have a normal distribution (See Appendix 4-D).

Data transformation involves the performance of a mathematical operation on each observation in the data set and the use of the transformed values in the statistical analysis. There are a number of transformations that can be used. A common transformation is the Log transformation**,** which consists of taking the log of each observation. To be consistent with engineering projects reviewed in the literature, the base-10 log transformation was used in the statistical test. Additionally, base-10 logs make it possible to look at the transformed number and see the magnitude of the original number. The orders of ten also relate to denominations of money (cost data), and the transformed data is easy to read. After log-transformation, the data has a much more normal distribution, as presented in Figure 4.8. The cost variables used in the data transformation are presented in Table 4.7.

Figure 4.8: Sample Log-Transformation Data Distribution, Cost per Mile



4.3.4 Final Variable Selection

Based on the statistical analyses of the data and variable relationships, a final assessment was performed to identify which variables would be the best predictors for modeling the cost per mile. When using a least squares regression procedure, the total number of predictor variables needs to be less than the number of observations. Analysis of the variables was performed to review the correlation of the predictor variables so that highly correlated variables were not used in the model development testing. In addition to the statistical analyses, the four basic criteria described previously (Section 4.1) were used to reassess the variables and to determine the final selection the variables to be used for model development and testing. The criteria are as follows: 1) Did the variable have sufficient data for each of the LRT systems? 2) Could the variable data potentially be quantified during the planning phase of project development? 3) Is the variable data mutually exclusive? 4) Could the data from multiple variables (sub-categories) be aggregated into a single variable?

Table 4.7: List of Variable Transformations – base-10 Log

Variable Name	Unit	SCC Category	SCC Definitions
Project_Total_Cost			
Cost_per_Mile			
Guideway_01_$	2015 Dollars	10.01	Guideway: At-grade exclusive right-of-way
Guideway_02_$	2015 Dollars	10.02	Guideway: At-grade semi-exclusive (allows cross-traffic)
Guideway_03_$	2015 Dollars	10.03	Guideway: At-grade in mixed traffic
Guideway_04_$	2015 Dollars	10.04	Guideway: Aerial structure
Guideway_05_$	2015 Dollars	10.05	Guideway: Built-up fill
Guideway_06_$	2015 Dollars	10.06	Guideway: Underground cut & cover
Guideway_07_$	2015 Dollars	10.07	Guideway: Underground tunnel
Guideway_08_$	2015 Dollars	10.08	Guideway: Retained cut or fill
Guideway_09_$	2015 Dollars	10.09	Track: Direct fixation
Guideway_10_$	2015 Dollars	10.10	Track: Embedded
Guideway_11_$	2015 Dollars	10.11	Track: Ballasted
Guideway_12_$	2015 Dollars	10.12	Track: Special (switches, turnouts)
Station_01_$	2015 Dollars	20.01	At-grade station, stop, shelter, mall, terminal, platform
Station_02_$	2015 Dollars	20.02	Aerial station, stop, shelter, mall, terminal, platform
Station_03_$	2015 Dollars	20.03	Underground station, stop, shelter, mall, terminal, platform
Station_06_$	2015 Dollars	20.06	Automobile parking multi-story structure
Support_Fac_$	2015 Dollars	30.01-.05	30 SUPPORT FACILITIES: YARDS, SHOPS, ADMIN. BLDGS
Sitework_$	2015 Dollars	40.01, .05, .08	Demolition, Clearing, Earthwork; Site structures including retaining walls, sound walls; Temporary Facilities and other indirect costs during construction
Utilities_$	2015 Dollars	40.02	Site Utilities, Utility Relocation
Environmental_$	2015 Dollars	40.03 - .04	Haz. mat'l, contam'd soil removal/mitigation, ground water treatments; Environmental mitigation, e.g. wetlands, historic/archeologic, parks
Accessways_$	2015 Dollars	40.06 - .07	
Systems_$	2015 Dollars		
ROW_$	2015 Dollars	60.01	Purchase or lease of real estate (60.02: Relocations)
Vehicles_$	2015 Dollars	70.01	Light Rail (70.6:Non-revenue vehicles; 70.07: spare parts)
Prof_Serv_$	2015 Dollars	80.01 - .08	Project Development

Utilizing these criteria and the statistical analyses, the final variable selection process resulted in 21 potential predictor variables. The selected variables and definitions are in Table 4.8. It should be noted that these are the variables to be considered and that correlated variables (as defined previously) were not included in the same model. For example, even though at grade guideway cost and length are variables to be tested, only one was examined at a time in the model development.

Table 4.8: Final Variable Selection

Variable #	Variable Name	Unit	Notes
1	Alignment_Length	Linear Feet	
2	Project_Total_Cost	2015 Dollars	
3	Cost_per_Mile	2015 Dollars	Response Variable
4	At_Grade_GW_Q	LF Guideway	Elements (10.01, 10.02, 10.03) Quantities
5	At_Grade_GW_$	2015 Dollars	Elements (10.01, 10.02, 10.03) Dollars
6	Aerial_GW_Q	LF Guideway	Elements (10.04 & 10.05) Quantities
7	Aerial_GW_$	2015 Dollars	Elements (10.04 & 10.05) Dollars
8	Below_GW_Q	LF Guideway	Elements (10.06, 10.07, 10.08) Quantities
9	Below_GW_$	2015 Dollars	Elements (10.06, 10.07, 10.08) Dollars
10	Station_01_Q	Stations	Quantity
11	Station_01_$	2015 Dollars	Revenue opening year $ converted to 2015
12	Station_02_Q	Stations	Quantity
13	Station_02_$	2015 Dollars	Revenue opening year $ converted to 2015
14	Station_03_Q	Stations	Quantity
15	Station_03_$	2015 Dollars	Revenue opening year $ converted to 2015
16	Station_Total_Q	Stations	Combines All Stations - Quantities
17	Station_Total_$	2015 Dollars	Combines All Stations converted to 2015 $
18	Utilities_$	2015 Dollars	SCC - 40.02
19	Environmental_$	2015 Dollars	Combine 40.03 + 40.04
20	ROW_Q	LF Guideway	Aggregate of all SCC (60.01 - 60.04) Quantity
21	ROW_$	2015 Dollars	Aggregate of all SCC (60.01 - 60.04) Dollars
22	Vehicles_Q	Vehicles	Aggregate of SCC 70.01, 70.06 & 70.07

4.3.5 Regression Analysis

The application of the multiple linear regression analysis was performed using the JMP software application of SAS. Since all variables in the model development data set were numeric and continuous, the least square procedure was applied.

Multiple linear regression techniques can be used to assess whether one continuous response variable (cost per mile) can be estimated from a group of predictor variables. A selection model procedure was utilized to reduce the number of predictor variables explaining the variance of the total group of predictors. The step-wise selection procedure was primarily used in developing the model. This procedure is a variation of both the forward selection and backwards elimination methods. It includes analysis at each step to determine the contribution of the predictor variable entered into the equation. This method allows for the determination of the contribution of each variable in the model and the additive statistical significance to the model of each variable used.

Through the use of the step-wise selection procedure, the variables from the final variable selection described above, which did not significantly contribute to the model were eliminated from additional tests. The *p*-values, or significance level, were examined to determine the variables to be eliminated. Variables with *p*-values less than 0.05 are considered to have significant contribution to the model. The actual linear multiple regression equation will take the following form:

$$Log(\hat{Y}) = a + b_1X_1 + b_2X_2 + \dots + b_nX_n$$

Where

$\hat{Y}$ = the predicated cost per mile of the response variable
a = the intercept constant
b_k = the multiple regression coefficient for the kth predictor
X_k = the kth predictor variable

The model development test process and final multiple linear regression model results are presented in the next section.

4.4 Model Results and Interpretations

The multiple regression technique was used for the model development, based on its flexible procedures, to analyze the single numeric continuous response variable of Cost per Mile and numeric continuous predictor variables.

The multiple regressions provided several criteria for determining the best-fit model to predict cost per mile for LRT systems at the planning level. Some of this information included: the determination of a significant relationship between cost per mile and the linear combination of predictor variables in each test equation; the review of *p*-value coefficients for each predictor variable to determine which coefficients were statistically significant; and the uniqueness of each predictor variable to determine those that account for a significant amount of variation in the cost per mile (response variable). In summary, the multiple regression procedure estimated the regression coefficients for the predictor variables, calculated R-squared, and tested for significance in order to establish the final model.

To validate the prediction performance of the regression analysis, one randomly selected project was removed from the data set. The regression analysis was performed using the 26 remaining LRT systems.

Several iterations with different variables were conducted to establish the best fit model. The parameter estimates in Table 4.9 are the terms that were included in the multiple regression equation from one of the better models during the early model development test process.

Table 4.9: Parameter Estimates – Preliminary Model

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	7.9595613	0.093234	85.37	<.0001*
At_Grade_GW_Q	-3.512e-6	1.291e-6	-2.72	0.0122*
Station_03_Q	0.152923	0.05895	2.59	0.0162*

The best predictor variables for this preliminary model were the length of At-Grade Guideway and the number of Underground Stations. This model produced an R-squared of 0.46. Through the use of the step-wise selection procedure (both forward selection and backwards elimination), several models were investigated for the inclusion of additional predictor variables. A model with additional variables having a *p*- value less than 0.05, which allows the rejection of the null hypothesis and tentatively concludes that the coefficient for the variable is significantly different from zero, was not identified at this stage of the model development process. Several "forced" models were attempted to include potential variables (e.g., at-grade stations, aerial stations and guideway, environmental cost, etc.) that would logically seem to be good predictors of Cost per Mile, but none were found to be significant. Based on previous experience in developing preliminary capital cost estimates, the SCC categories of Guideway and Stations seem that they would be the best predictors at the planning level. However, it was unexpected that the early model development trials did not include some of these sub-category elements as predictors, so additional model development investigation and tests were performed.

As previously stated, several model iterations with different variables were conducted to establish the best fit model. Many of the early model runs were not significant beyond a single predictor variable. In attempts to develop a better fit model with additional predictor variables, and in order to create stronger linear relationships between the response and predictor variables, many of the variables were examined for additional transformation to be included in further model development test. This was done by identifying the variables that gave the best fit in the previous model tests.

The nonlinear transformations were performed to increase the linear relationship between the response and predictor variables in anticipation of creating a stronger correlation between the variables. There are many methods to transform variables to achieve stronger linearity. The chosen method needs to be tested on the data to check if it increases the linear relationship, on a trial and error basis. An exponential transformation method was used in this research. The variables were transformed by an exponential process of either taking the square of the variable, taking the variable to the power of 0.5, or both. After the predictor variables were exponentially transformed, additional regression analyses were performed which computed the coefficient of determination (R-squared) and the variables corresponding *p*- value.

Several more models were attempted using a combination of both the transformed and raw data variables. Some of the potential variables that would logically seem to be good predictors of Cost per Mile included: at-grade stations, aerial stations and guideway, total guideway costs, total station cost, utilities cost, and environmental cost, among others. The regression analysis was conducted to compute the R-squared values of the regression models using the transformed variables and some stronger models were developed. Many of the better-fit models included cost variables (e.g., total guideway cost, total station costs) that can be very difficult to estimate during early planning phases due to the lack of available engineering data. The parameter estimates in Table 4.10 are the terms that make up the multiple regression equation from one of the better models with cost variables.

Table 4.10: Parameter Estimates – Candidate Model with Cost Variables

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	6.625195	0.439516	15.07	<.0001*
Alignment^.5	-0.002777	0.000528	-5.26	<.0001*
Total_GW_$_Squared	0.0242817	0.005986	4.06	0.0005*
Station_Total_$_Squared	0.0049269	0.002258	2.18	0.0401*

This model produced an R-squared of 0.66 and significant p values. However, since the cost variables that are required for this model would be very difficult for a Transportation Planner to estimate during conceptual planning phases, this model was not selected as the final model.

The best fit model developed, which includes variables that could be expected to be developed during early transit corridor planning, was selected as the final model. The variables include alignment length in linear feet (Alignment_Length) and underground station quantity (Station_03_Q). Both these predictor variables were transformed exponentially by the power of 0.5. The parameter estimates in Table 4.11 are the terms that make up the multiple regression equation for the final model. The final multiple linear regression model equation is as follows:

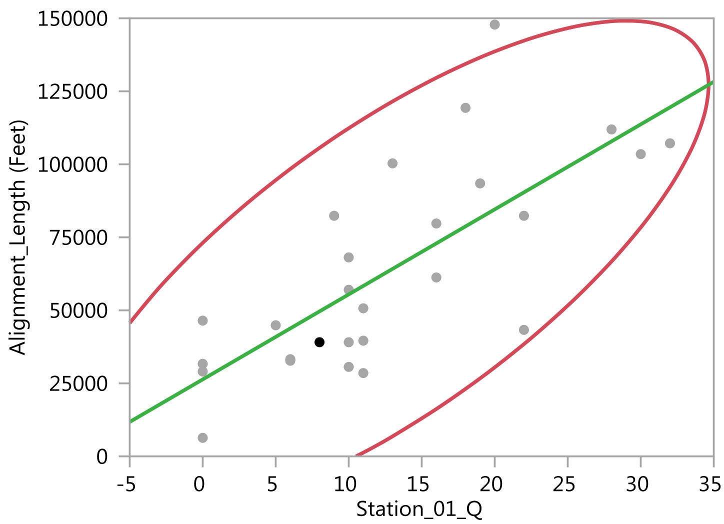
$$\textbf{Log (Cost per mile)} = 8.2719299 - 0.00219\,(\text{Alignment\_Length})^{0.5} + 0.2533808(\text{Station\_03\_Q})^{0.5}$$

Table 4.11 - Parameter Estimates – Final Model

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	8.2719299	0.145926	56.69	<.0001*
Alignment^.5	-0.00219	0.000565	-3.87	0.0008*
Station_03_Q^.5	0.2533808	0.071932	3.52	0.0018*

While the total route alignment length seems logical as a strong predictor of LRT cost, it was unanticipated that the underground stations would be a strong predictor because the LRT stations for the majority of the systems were at-grade. At-grade and aerial stations performed better than many of the other potential model variables, but not at an acceptable level for inclusion in the final model. However, at-grade stations can be assumed to be accounted for in total alignment length. The correlation analyses of the predictor variables previously discussed indicated that the alignment length and at-grade stations were highly correlated variables. Highly correlated predictor variables should not be used in the model at the same time to minimize the problems that may occur with multicollinearity. Figure 4.9 shows the scatterplot with a 95% density ellipse overlaid on the pairwise correlation of the Alignment_Length and Station_01_Q predictor variables. The Pearson correlation coefficient between the two variables is 0.74 and the calculated p-value is less than 0.0001, which indicates a very strong correlation. The scatterplot also shows the bivariate fit of the two variables and indicates a positive linear relationship; as the alignment length increases so does the quantity of at-grade stations.

Figure 4.9: Correlation and Bivariate Fit of Alignment_Length (Feet) By Station_01_Q

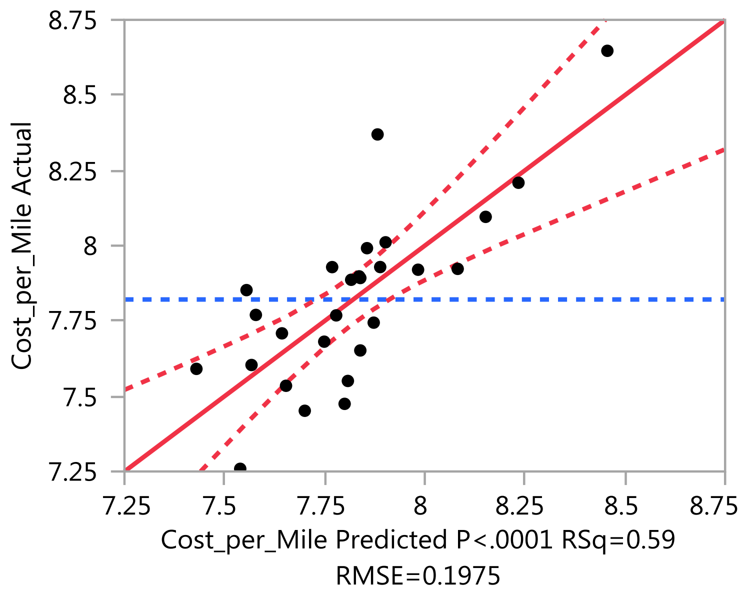


This model produced an R-squared of 0.59 and a very small p-value (<.0001). This R-squared indicates that 59 percent of the variance in cost per mile (response) is accounted for by the linear combination of the predictor variables.

The transformed variable model R-squared (0.59) is greater than the R-squared of the raw score R-squared (0.51) when tested on the same non-transformed variables.

Review of the Actual by Predicted Plot (Figure 4.10) indicates that the multiple regression model is statistically significant and that the linear combination of the two predictor variables accounts for a significant amount of variation in cost per mile. In the plot, the fitted and confidence lines cross the horizontal reference line, which confirms a significant regression model.

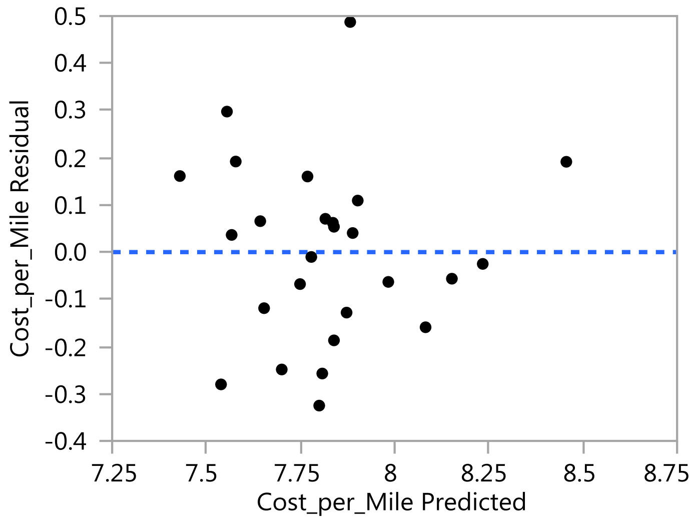
Figure 4.10: Actual by Predicted Plot



In the Analysis of Variance statistics calculated by the final model run, the *F* value produced was 16.8 and its associated probability was less than 0.0001. The small *p*-value allows for rejection of the null hypothesis (that all effects are zero) and assumes that the predictor variables account for a significant amount of the variation in cost per mile (response variable).

In order to validate model assumptions, a residual analysis can be conducted by examination of the residuals. The model's residual plot is presented in Figure 4.11. The plot illustrates the residuals on the vertical axis and the response variable (cost per mile) on the horizontal axis. The residual is the difference between the observed (actual) value of the response variable (Y) and the predicted value ($\hat{Y}$). Review of the plot indicates that the linear regression model is appropriate because the data points are randomly dispersed around the horizontal axis. If they were not, a non-linear model would be more appropriate.

Figure 4.11: Residual by Predicted Plot



4.5 Statistical Validation of Cost Estimating Results

Cross validation is a method measuring the accuracy and assessing the predictive ability of a statistical model by testing the model on data not used in the model estimation. The data not used in the model development can be referred to as the test set. The data used in developing the model is the training set. Once the optimal statistical model is selected based on the training set data, one can check the model's predictive capability on the test set.

For the purpose of validating the regression model developed for this research, one LRT system was randomly selected using Excel's random number generator function and then removed from the data set. The predicted cost per mile was calculated using the final multiple linear regression model equation described above. The project removed was Project 18, which had a predicted cost per mile of $71,748,880 and actual cost per mile of $98,094,705, resulting in a 26.9% error. At first glance, the percentage error seems large, but as a predictor at the planning level, the model has produced an acceptable cost estimate, since a 30% contingency (or higher) is typically applied to early level cost estimates.

Additionally, a comparison was made for the other 26 LRT systems used in the model estimation and is presented in Table 4.12. The percent error range is from 2.4% to 111.5%. A bivariate fit analysis was performed using the LRT systems' actual cost per mile by the model-predicted cost per mile and is presented in Figure 4.12. The fit is linear and illustrates that the model predictions of cost per mile are fairly accurate for approximately 60% of the LRT projects. The model predicted 13 projects with a higher cost per mile and 14 projects with a lower cost per mile than the actual cost. The actual cost per mile by the model-predicted cost per mile in 2015 dollars is presented graphically as a bar chart in Figure 4.13.

Figure 4.12: Bivariate Fit of Cost_per_Mile by Predicted Cost per Mile

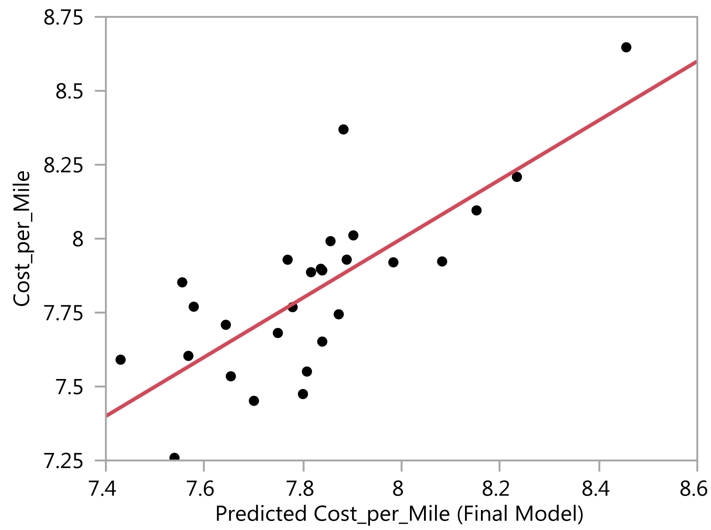
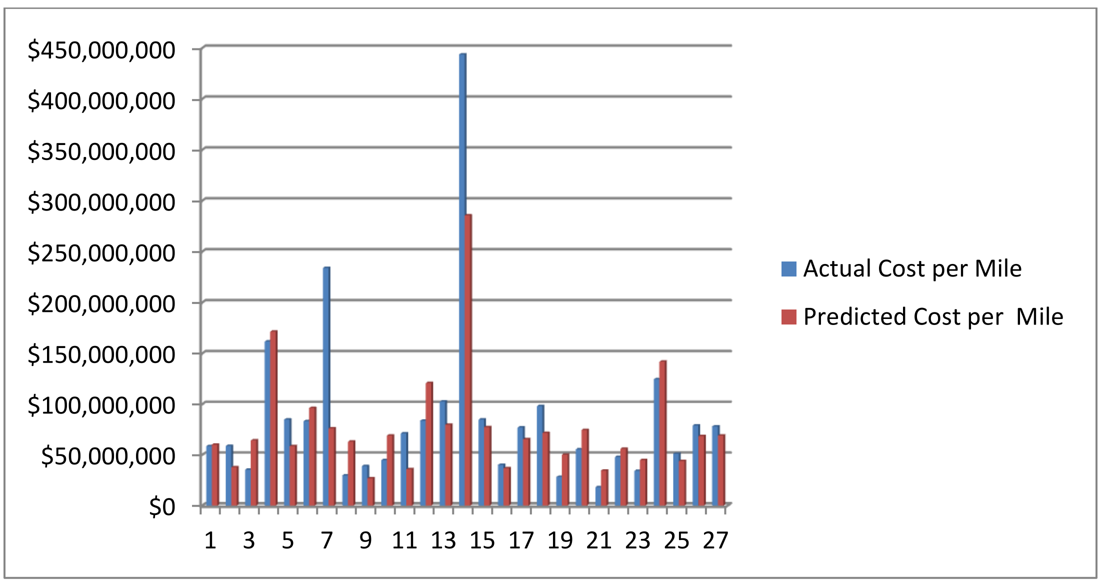


Figure 4.13: Actual Compared to Predicted Cost per Mile



2015 Dollars

Additionally, the removed project used in the cross validation was run with the CCD cost model. The results produced a 33% error which was greater than this research's model. As described previously in Section 2.3 (Predicted and Actual Cost of Major Transit Systems), FTA has conducted three studies (1990, 2003 and 2007) that analyzed the predicted actual capital cost and ridership of major transit projects that were constructed utilizing federal funds. The research model performed adequately (average of 35% error) compared to the predicted and actual cost of major built transit systems in these studies (see Table 2.2).

Table 4.12: Actual and Predicted Cost per Mile

Project #*	Actual Cost per Mile (Log)	Predicted Cost per Mile (Log)	Actual Cost per Mile	Predicted Cost per Mile	Percent Error
1	7.768448854	7.778873395	$58,674,427	$60,099,851	2.43%
2	7.769988717	7.578283915	$58,882,836	$37,869,007	35.69%
3	7.55066201	7.807980584	$35,535,466	$64,265,899	80.85%
4	8.208926837	8.234025736	$161,780,747	$171,405,888	5.95%
5	7.92880174	7.768799616	$84,879,290	$58,721,835	30.82%
6	7.920051911	7.983322201	$83,186,320	$96,232,596	15.68%
7	8.369426123	7.882134508	$234,113,320	$76,231,507	67.44%
8	7.474604923	7.799864237	$29,826,681	$63,076,013	111.48%
9	7.590784887	7.429875612	$38,974,889	$26,907,640	30.96%
10	7.651970513	7.839040516	$44,871,492	$69,030,420	53.84%
11	7.852405592	7.554946509	$71,187,803	$35,887,773	49.59%
12	7.922930811	8.082272135	$83,739,586	$120,857,090	44.32%
13	8.011202611	7.902137521	$102,613,054	$79,824,742	22.21%
14	8.647176695	8.455942665	$443,789,165	$285,721,331	35.62%
15	7.928996737	7.888686157	$84,917,409	$77,390,234	8.86%
16	7.603648806	7.567416737	$40,146,603	$36,933,183	8.00%
17	7.886656087	7.81624147	$77,029,324	$65,500,026	14.97%
18	7.991645564	7.855815127	$98,094,705	$71,748,880	26.86%
19	7.451597312	7.70037766	$28,287,679	$50,162,325	77.33%
20	7.744276687	7.87250848	$55,497,917	$74,560,443	34.35%
21	7.258834376	7.539225151	$18,148,234	$34,611,877	90.72%
22	7.681100216	7.748964503	$47,984,416	$56,100,212	16.91%
23	7.534644657	7.65355828	$34,248,745	$45,035,841	31.50%
24	8.095669482	8.152110025	$124,643,456	$141,941,707	13.88%
25	7.708780497	7.643403715	$51,142,328	$43,995,040	13.98%
26	7.898105744	7.836125403	$79,087,117	$68,568,619	13.30%
27	7.892614869	7.839040516	$78,093,497	$69,030,420	11.61%

** Project # corresponds to the project name in Table 3.3*

In addition to the cross validation analysis performed for the randomly selected LRT system (Project 18) removed from the data set to develop and select the final model equation, an analysis was conducted on the complete project list (Projects 1 through 27) using the final model's predictor variables. Each of the 26 remaining LRT projects was removed from the data set, one project at a time, to generate model runs using the alignment length and underground station as predictor variables. Each of the 26 model runs produced a set of new model coefficients during the model development. Next, the cost per mile for the removed project (test set data) was estimated with its respective model to check the predictive capability of the model.

The results of this analysis were very similar to the initial cross validation analysis using Project 18. The predicted cost per mile was calculated using the multiple linear regression model equation for each of the additional 26 models developed. This resulted in a percent error range from 2.5% to 110.8%, with approximately half the projects with a 30% or lower percent error.

A comparison was made for the other 26 LRT systems used in the model estimation on each of the model equations developed in the expanded cross validation. This analysis also produced very similar results to the initial cross validation analysis, with an average percent error for all projects ranging from 34% to 37%. The model coefficients, actual and predicted cost per mile, percent error for the removed (test set) project, and the average percent error across all 27 projects for each of the cross validation models are presented in Table 4.13.

In summary, the model appears to be a useful tool for estimating LRT cost per mile at the planning level when only limited alignment data is available. However, further development of better predictive models will become possible when additional LRT system data becomes available.

Table 4.13: Expanded Cross Validation Results - Actual and Predicted Cost per Mile

Project #*	Model Coefficients			Actual Cost/Mile	Predicted Cost/Mile	Project % Error	% Error for all Projects	% Error - Low	% Error - High
	Intercept	Alignment^.5	Station_03_Q^.5						
1	8.2547288	-0.002113	0.2654548	$58,674,427	$60,118,491	2.46%	35.57%	2.46%	111.18%
2	8.2709805	-0.002221	0.2718356	$58,882,836	$36,941,607	37.26%	35.22%	0.58%	107.79%
3	8.2855547	-0.002185	0.2542652	$35,535,466	$66,475,962	87.07%	36.27%	4.72%	118.76%
4	8.2575271	-0.002127	0.2720426	$161,780,747	$180,889,499	11.81%	35.78%	2.38%	111.07%
5	8.2899232	-0.002265	0.2491861	$84,879,290	$57,107,126	32.72%	35.10%	2.69%	112.37%
6	8.2519233	-0.0021	0.2731082	$83,186,320	$101,231,693	21.69%	35.85%	2.49%	111.18%
7	8.2914972	-0.002195	0.251358	$234,113,320	$79,581,453	66.01%	36.45%	4.12%	120.67%
8	8.1557898	-0.001821	0.2912127	$29,826,681	$57,978,098	94.38%	34.41%	5.08%	94.38%
9	8.2936480	-0.002309	0.2697229	$38,974,889	$25,458,819	34.68%	35.36%	1.24%	109.57%
10	8.2823266	-0.002187	0.2570648	$44,871,492	$70,799,504	57.78%	36.09%	5.07%	116.92%
11	8.2877340	-0.002312	0.2749011	$71,187,803	$33,947,599	52.31%	35.09%	0.28%	106.43%
12	8.2134716	-0.001967	0.3145162	$83,739,586	$156,208,285	86.54%	36.78%	0.50%	106.48%
13	8.2285721	-0.002034	0.2719746	$102,613,054	$76,757,549	25.20%	35.17%	0.51%	106.79%
14	8.1765429	-0.001793	0.2336426	$443,789,165	$231,335,581	47.87%	34.41%	0.90%	107.89%
15	8.2445438	-0.002083	0.2682184	$84,917,409	$75,862,081	10.66%	35.41%	1.65%	109.38%
16	8.2567782	-0.002129	0.2668079	$40,146,603	$37,315,431	7.05%	35.50%	2.10%	110.50%
17	8.2442177	-0.002088	0.2692124	$77,029,324	$64,528,729	16.23%	35.34%	1.32%	108.71%
18	8.2719299	-0.00219	0.2533808	$98,094,705	$71,748,880	26.86%	35.15%	2.43%	111.48%
19	8.2596624	-0.002082	0.256235	$28,287,679	$52,035,089	83.95%	36.12%	3.12%	116.91%
20	8.2773062	-0.002179	0.2595842	$55,497,917	$75,838,722	36.65%	35.94%	4.30%	115.28%
21	8.2150694	-0.00189	0.2571159	$18,148,234	$38,259,137	110.81%	36.05%	0.79%	115.31%
22	8.2584521	-0.002116	0.2630536	$47,984,416	$56,644,491	18.05%	35.71%	3.18%	112.68%
23	8.2512906	-0.002075	0.2614768	$34,248,745	$46,279,701	35.13%	35.79%	3.68%	113.51%
24	8.2630451	-0.002145	0.271358	$124,643,456	$147,526,803	18.36%	35.86%	2.72%	111.87%
25	8.2555610	-0.002131	0.2680413	$51,142,328	$44,052,211	13.86%	35.42%	1.70%	109.70%
26	8.2439746	-0.002085	0.2689567	$79,087,117	$67,462,874	14.70%	35.36%	1.42%	108.90%
27	8.2450718	-0.002088	0.2685748	$78,093,497	$67,974,244	12.96%	35.38%	1.52%	109.12%

** Project # corresponds to the project name in Table 3.3*

5. CONCLUSIONS AND FUTURE STEPS

This chapter provides conclusions regarding the development and results of the LRT capital cost estimating model, identifies the research contributions to the current state of the practice, and suggests steps for future research to advance the model's predictive accuracy.

5.1 Summary of Work Completed

The primary problem addressed in this research is associated with producing capital cost estimates at the planning level while considering LRT as an alternative mode of public transportation in the study corridor. The capital cost estimates for each mode of public transportation under study must be sensitive to a range of independent variables, such as vertical and horizontal alignment characteristics, environmentally sensitive areas, urban design, and other unique cost-controlling factors.

A comprehensive literature review summarizing the state-of-the-art in transit project evaluation was undertaken with an emphasis on capital cost estimating, along with current methodologies employed and limitations encountered in current practice. In this research, statistical theory is utilized to enhance the reliability of developing LRT cost estimates used for improving alternative transportation system development decisions.

A wealth of data has been gathered defining the basis for engineering decision-making at the System Planning level. This research identifies the effective determinants of project cost for LRT projects and suggests the use of these determinants at the earliest level of system planning. An important source of this data is the information required by FTA and the NEPA-mandated major capital project development process.

Model development activities include sample size selection, model framework and selection, and model development and testing. The developed model utilizes statistical theory to enhance the quality of capital cost estimation for LRT investments by varying alignment characteristics.

For the purpose of validating the regression model developed for this research, one LRT system was removed from the data set and run through the final multiple linear regression model equation to assess the model's predictive accuracy. Comparing the model's estimated cost to the projects final construction cost resulted in a 26.9% error. The percentage error seems somewhat high, but acceptable at the planning level since a 30% contingency (or higher) is typically applied to early level cost estimates.

Additionally, a comparison was made for the other 26 LRT systems used in the model estimation. The percent error range was from 2.4% to 111.5%, with just over 60% of the 26 projects' predicted cost estimate within 30% or better of their actual cost. The model appears to be a useful tool for estimating LRT cost per mile at the planning level when only very limited alignment data is available. However, there is much room for improvement to develop a better predictive model once additional LRT system data becomes available.

5.2 Contribution of this Research to State of the Practice

The primary problem addressed in this research is the challenge associated with producing capital cost estimates at the planning level for the LRT mode of public transportation in the study corridor. Capital cost estimates are an important element in calculating the cost-effectiveness, financial requirements and implementation feasibility of major capital transit investments. Cost plays a critically important role in the evaluation of alternatives and the decision-making process. The goal of this dissertation was to develop an improved, simplified tool for use at the planning level for LRT capital cost estimates as a component of the evaluation process.

The review of the literature reflects a wide range of estimates of capital cost for the LRT mode of public transportation. Currently, deficient methodologies rely on either LRT cost averages or high and low cost ranges without regard to specific project alignment characteristics and those that cannot produce accurate estimates due to lack of engineering data at the planning level of project development. Most of these estimates are based on a review of cost data of implemented LRT projects; these estimates lack any systematic approach that can be followed in future evaluations of alternatives. This dissertation research provides a useful tool to transportation planners. This tool is useful because it can be used for developing reliable LRT capital cost estimates at the corridor level to be used in the early stages of project development by the provision of a statistically significant model to estimate planning level cost.

The current State of the Practice for the development of conceptual or planning level transit project capital cost estimates includes use of the FTA Capital Cost Database. The CCD is a Microsoft Access database of as-built costs for 54 federally-funded projects in the following modes: Bus Rapid Transit, Commuter Rail, Light Rail, Heavy Rail, and Trolley. The purpose of the CCD is developing conceptual, "order-of-magnitude" cost estimates for potential projects in the modes listed above.

This research takes a different approach than the CCD methodology through its application of statistical techniques for predicting LRT cost at the corridor level rather than use of an Access database. The models developed herein can be embedded in the current cost-effectiveness evaluation techniques used in the current state of practice. Engineering decision-making can be enhanced because of this improved, simplified cost-estimating methodology, and reflects advancement in transit corridor evaluation and the overall state of the practice of transit systems planning.

Specific research contributions that could be applied elsewhere are summarized below:

- Development of a comprehensive data set that includes detailed capital cost data for 27 implemented LRT systems. The data set could be useful for multiple transit planning purposes and for other transit modes (HRT, BRT) that have similar alignment characteristics.

- In addition to the use of the multiple regression models to estimate LRT planning level capital cost, the data set could be used in various spreadsheet applications regarding corridor level transit planning cost analyses, similar to the planning cost estimating methodologies described in the literature (based on LRT cost averages of SCC categories per length of route alignment and ranges of high and low cost categories).

- Current sources that sell and publish cost estimating data typically do not make information on the development of the data and estimates available. The industry needs reliable and replicable procedures for estimating modal capital cost during the planning phases of major capital transit project development.

- All major metropolitan areas would benefit from a crisply defined and successfully tested capital cost estimating methodology for establishing the investment-worthiness of LRT alternatives for solving their mobility needs.

- Development of a simplified model requiring minimal alignment data details to generate planning level LRT capital cost estimates that could be replicated in other major metropolitan areas.

- Improvement on widely varied, currently available, estimates of LRT capital cost for public transportation modal options.

5.3 Recommendations for Further Research

One method to build on this research would be the collection of additional verified capital cost data from new LRT systems as they are implemented in the near future and the cost data is made available for distribution by FTA. The data set was limited to 27 LRT systems and a sample size of 100 or more is desirable in order to improve the model's predictive capability. This current research could be continued by collecting cost data as more LRT projects are completed and the data becomes available. Additionally, data from recently implemented LRT systems outside the U.S. could be collected and potentially used to continue to pursue this research.

Another area where this research could be extended is through the introduction and examination of locational factors regarding cost variables. A majority of costs in construction are local costs, causing cost to vary from place to place due to local cost of living (Hoback, 2008). Differences in cost for the same LRT system element from project to project could include the impact of unions and labor cost within the construction sector, as well as the costs of materials and transportation factors.

The research could also be advanced by the investigation of more highly sophisticated statistical techniques beyond multiple regression analysis. Factor analysis and/or principal component analysis is a way to reduce the number of variables to a smaller number of uncorrelated variables that account for most of the variance in the data set. Although this dissertation did not utilize these advanced statistical techniques, they are worthy of future investigation for producing accurate cost estimates at early level transportation project planning phases.

APPENDIX

Appendix 3-A: Sample SCC Worksheet for a New Start Build Alternative

MAIN WORKSHEET-BUILD ALTERNATIVE		(Rev.17, June, 2015)
Insert Project Sponsor's Name here	Today's Date	**6/22/15**
Insert Project Name and Location	Yr of Base Year $	2015
Insert Current Phase (e.g. Applic. for Engineering, Engineering, Applic. for FFGA, Construction, Rev Ops)	Yr of Revenue Ops	2020

For all cells, enter costs to the $1. Note, all costs will be displayed to the nearest $1,000!

	Quantity	Base Year Dollars w/o Contingency (X000)	Base Year Dollars Allocated Contingency (X000)	Base Year Dollars TOTAL (X000)	Base Year Dollars Unit Cost (X000)	Base Year Dollars Percentage of Construction Cost	Base Year Dollars Percentage of Total Project Cost	YOE Dollars Total (X000)
10 GUIDEWAY & TRACK ELEMENTS (route miles)	**10.00**	80,000	20,000	**100,000**	**$10,000**	**45%**	**25%**	**104,780**
10.01 Guideway: At-grade exclusive right-of-way	10.00	80,000	20,000	100,000	$10,000			104,780
10.02 Guideway: At-grade semi-exclusive (allows cross-traffic)				0				0
10.03 Guideway: At-grade in mixed traffic				0				0
10.04 Guideway: Aerial structure				0				0
10.05 Guideway: Built-up fill				0				0
10.06 Guideway: Underground cut & cover				0				0
10.07 Guideway: Underground tunnel				0				0
10.08 Guideway: Retained cut or fill				0				0
10.09 Track: Direct fixation				0				0
10.10 Track: Embedded				0				0
10.11 Track: Ballasted				0				0
10.12 Track: Special (switches, turnouts)				0				0
10.13 Track: Vibration and noise dampening				0				0
20 STATIONS, STOPS, TERMINALS, INTERMODAL (number)	**20**	28,000	2,000	**30,000**	**$1,500**	**13%**	**8%**	**32,149**
20.01 At-grade station, stop, shelter, mall, terminal, platform	20	28,000	2,000	30,000	$1,500			32,149
20.02 Aerial station, stop, shelter, mall, terminal, platform				0				0
20.03 Underground station, stop, shelter, mall, terminal, platform				0				0
20.04 Other stations, landings, terminals: Intermodal, ferry, trolley, etc.				0				0
20.05 Joint development				0				0
20.06 Automobile parking multi-story structure				0				0
20.07 Elevators, escalators				0				0
30 SUPPORT FACILITIES: YARDS, SHOPS, ADMIN. BLDGS	10.00	8,000	2,000	**10,000**	**$1,000**	**4%**	**3%**	**10,531**
30.01 Administration Building: Office, sales, storage, revenue counting				0				0
30.02 Light Maintenance Facility		8,000	2,000	10,000				10,531
30.03 Heavy Maintenance Facility				0				0
30.04 Storage or Maintenance of Way Building				0				0
30.05 Yard and Yard Track				0				0
40 SITEWORK & SPECIAL CONDITIONS	10.00	44,350	9,000	**53,350**	**$5,335**	**24%**	**14%**	**54,167**
40.01 Demolition, Clearing, Earthwork		9,000	1,800	10,800				10,965
40.02 Site Utilities, Utility Relocation		18,000	3,600	21,600				21,931
40.03 Haz. mat'l, contam'd soil removal/mitigation, ground water treatments				0				0
40.04 Environmental mitigation, e.g. wetlands, historic/archeologic, parks		9,000	1,800	10,800				10,965
40.05 Site structures including retaining walls, sound walls				0				0
40.06 Pedestrian / bike access and accommodation, landscaping		8,350	1,800	10,150				10,305
40.07 Automobile, bus, van accessways including roads, parking lots				0				0
40.08 Temporary Facilities and other indirect costs during construction				0				0
50 SYSTEMS	10.00	25,000	5,000	**30,000**	**$3,000**	**13%**	**8%**	**32,149**
50.01 Train control and signals		9,000	1,000	10,000				10,716
50.02 Traffic signals and crossing protection		2,000	500	2,500				2,679
50.03 Traction power supply: substations		4,000	1,000	5,000				5,358
50.04 Traction power distribution: catenary and third rail		4,000	1,000	5,000				5,358
50.05 Communications		4,000	1,000	5,000				5,358
50.06 Fare collection system and equipment		2,000	500	2,500				2,679
50.07 Central Control				0				0
Construction Subtotal (10 - 50)	10.00	185,350	38,000	**223,350**	**$22,335**	**100%**	**57%**	**233,777**
60 ROW, LAND, EXISTING IMPROVEMENTS	10.00	30,175	5,000	**35,175**	**$3,518**		**9%**	**35,350**
60.01 Purchase or lease of real estate		30,175	5,000	35,175				35,350
60.02 Relocation of existing households and businesses				0				0
70 VEHICLES (number)	**15**	27,000	3,000	**30,000**	**$2,000**		**8%**	**31,593**
70.01 Light Rail	15	27,000	3,000	30,000	$2,000			31,593
70.02 Heavy Rail				0				0
70.03 Commuter Rail				0				0
70.04 Bus				0				0
70.05 Other				0				0
70.06 Non-revenue vehicles				0				0
70.07 Spare parts				0				0
80 PROFESSIONAL SERVICES (applies to Cats. 10-50)	10.00	70,000	7,168	**77,168**	**$7,717**	**35%**	**20%**	**77,424**
80.01 Project Development		5,000	668	5,668				5,687
80.02 Engineering		20,000	2,000	22,000				22,073
80.03 Project Management for Design and Construction		5,000	500	5,500				5,518
80.04 Construction Administration & Management		20,000	2,000	22,000				22,073
80.05 Professional Liability and other Non-Construction Insurance		5,000	500	5,500				5,518
80.06 Legal; Permits; Review Fees by other agencies, cities, etc.		5,000	500	5,500				5,518
80.07 Surveys, Testing, Investigation, Inspection		5,000	500	5,500				5,518
80.08 Start up		5,000	500	5,500				5,518
Subtotal (10 - 80)	10.00	312,525	53,168	**365,693**	**$36,569**		**93%**	**378,144**
90 UNALLOCATED CONTINGENCY				**20,000**			**5%**	**21,075**
Subtotal (10 - 90)	10.00			**385,693**	**$38,569**		**98%**	**399,219**
100 FINANCE CHARGES				**7,600**			**2%**	**9,500**
Total Project Cost (10 - 100)	10.00			**393,293**	**$39,329**		**100%**	**408,719**
Allocated Contingency as % of Base Yr Dollars w/o Contingency				17.01%				
Unallocated Contingency as % of Base Yr Dollars w/o Contingency				6.40%				
Total Contingency as % of Base Yr Dollars w/o Contingency				23.41%				
Unallocated Contingency as % of Subtotal (10 - 80)				5.47%				
YOE Construction Cost per Mile (X000)								$23,378
YOE Total Project Cost per Mile Not Including Vehicles (X000)								$37,713
YOE Total Project Cost per Mile (X000)								$40,872

Enter finance charges on Inflation Worksheet.

Appendix 4-A: Historical Cost Indices - *RSMeans Construction Cost*

Historical Cost Indexes

The table below lists both the RSMeans® historical cost index based on Jan. 1, 1993 = 100 as well as the computed value of an index based on Jan. 1, 2016 costs. Since the Jan. 1, 2016 figure is estimated, space is left to write in the actual index figures as they become available through the quarterly *RSMeans Construction Cost Indexes*.

To compute the actual index based on Jan. 1, 2016 = 100, divide the historical cost index for a particular year by the actual Jan. 1, 2016 construction cost index. Space has been left to advance the index figures as the year progresses.

Year	Historical Cost Index Jan. 1, 1993 = 100		Current Index Based on Jan. 1, 2016 = 100		Year	Historical Cost Index Jan. 1, 1993 = 100	Current Index Based on Jan. 1, 2016 = 100		Year	Historical Cost Index Jan. 1, 1993 = 100	Current Index Based on Jan. 1, 2016 = 100	
	Est.	Actual	Est.	Actual		Actual	Est.	Actual		Actual	Est.	Actual
Oct 2016*					July 2001	125.1	60.4		July 1983	80.2	38.7	
July 2016*					2000	120.9	58.3		1982	76.1	36.8	
April 2016*					1999	117.6	56.8		1981	70.0	33.8	
Jan 2016*	207.2		100.0	100.0	1998	115.1	55.6		1980	62.9	30.4	
July 2015		206.2			1997	112.8	54.4		1979	57.8	27.9	
2014		204.9	98.9		1996	110.2	53.2		1978	53.5	25.8	
2013		201.2	97.1		1995	107.6	51.9		1977	49.5	23.9	
2012		194.6	93.9		1994	104.4	50.4		1976	46.9	22.6	
2011		191.2	92.3		1993	101.7	49.1		1975	44.8	21.6	
2010		183.5	88.6		1992	99.4	48.0		1974	41.4	20.0	
2009		180.1	86.9		1991	96.8	46.7		1973	37.7	18.2	
2008		180.4	87.1		1990	94.3	45.5		1972	34.8	16.8	
2007		169.4	81.8		1989	92.1	44.5		1971	32.1	15.5	
2006		162.0	78.2		1988	89.9	43.4		1970	28.7	13.9	
2005		151.6	73.2		1987	87.7	42.3		1969	26.9	13.0	
2004		143.7	69.4		1986	84.2	40.7		1968	24.9	12.0	
2003		132.0	63.7		1985	82.6	39.9		1967	23.5	11.3	
2002		128.7	62.1		1984	82.0	39.6		1966	22.7	11.0	

Adjustments to Costs

The "Historical Cost Index" can be used to convert national average building costs at a particular time to the approximate building costs for some other time.

Example:

Estimate and compare construction costs for different years in the same city.

To estimate the national average construction cost of a building in 1970, knowing that it cost \$900,000 in 2016:

INDEX in 1970 = 28.7

INDEX in 2016 = 207.2

Time Adjustment Using the Historical Cost Indexes:

$$\frac{\text{Index for Year A}}{\text{Index for Year B}} \times \text{Cost in Year B} = \text{Cost in Year A}$$

$$\frac{\text{INDEX 1970}}{\text{INDEX 2016}} \times \text{Cost 2016} = \text{Cost 1970}$$

$$\frac{28.7}{207.2} \times \$900{,}000 = .139 \times \$900{,}000 = \$124{,}662$$

The construction cost of the building in 1970 was \$124,662.

Note: The city cost indexes for Canada can be used to convert U.S. national averages to local costs in Canadian dollars.

Example:

To estimate and compare the cost of a building in Toronto, ON in 2016 with the known cost of \$600,000 (US\$) in New York, NY in 2016:

INDEX Toronto = 109.9

INDEX New York = 131.1

$$\frac{\text{INDEX Toronto}}{\text{INDEX New York}} \times \text{Cost New York} = \text{Cost Toronto}$$

$$\frac{109.9}{131.1} \times \$600{,}000 = .841 \times \$600{,}000 = \$502{,}975$$

The construction cost of the building in Toronto is \$502,975 (CN\$).

*Historical Cost Index updates and other resources are provided on the following website:
http://info.thegordiangroup.com/RSMeans.html

Appendix 4-B: Potential Model Variables - Data Table

	Variable #:	1	2	3	4
Project #	Project Name	Alignment_Length	Project_Total_Cost	Cost_per_Mile	Guideway_01_Q
1	Charlotte - South Corridor LRT	9.6	462748292	48202947	13200
2	Denver - Southeast Corridor - T-Rex	19.0	878959094	46261005	100320
3	Denver Southwest Corridor	8.5	177100001	20835294	37149
4	Los Angeles - East Side Extension	6.2	876079616	141303164	0
5	Los Angeles - Long Beach Blue Line	22.6	877269855	38817250	63487
6	Minneapolis - Hiawatha Corridor	11.6	672477878	57972231	44248
7	New Jersey - Hudson-Bergen MOS-2	6.0	1006165000	167694167	31632.30584
8	New Jersey - Newark Rail Link	8.8	206212000	23433182	45839.30908
9	Southern New Jersey LRT	28.0	698599350	24949977	139957.5
10	Norfolk - Light Rail Transit	7.4	307894159	41607319	25872
11	Phoenix - Central Phoenix/East Valley	20.3	1264298130	62280696	104016
12	Pittsburgh Light Rail Stage I	15.6	555605245	35615721	34933
13	Pittsburgh Light Rail Stage II	5.4	386157500	71510648	7588
14	Pittsburgh - North Shore LRT Connector	1.2	502588000	418823333	0
15	Portland Interstate MAX	5.8	343236000	59178621	11193.6
16	Portland MAX Segment I	19.6	321313000	16393520	0
17	Portland - South Corridor/Portland Mall	8.2	551689839	67279249	36036
18	Portland/Westside/Hillsboro MAX	17.7	969182332	54756064	76137.6
19	Sacramento Folsom Corridor	12.9	268285714	20797342	59928
20	Sacramento South Corridor	6.3	223821859	35527279	32614
21	Sacramento Stage I	21.2	163636863	7718720	106920
22	Salt Lake City - Mid Jordan LRT	10.8	480532969	44493793	55968
23	Salt Lake North South Corridor	15.1	294944466	19532746	65208
24	San Diego Mission Valley East	5.5	504014126	91638932	0
25	Santa Clara VTA North Corridor	15.6	339325126	21751611	82252
26	Santa Clara VTA Tasman West	7.5	359861719	47981563	0
27	St Louis St Clair Cnty Extension	7.4	350602680	47378741	91872

	Variable #:	5	6	7	8
Project #	Project Name	Guideway_01_$	Guideway_02_Q	Guideway_02_$	Guideway_03_Q
1	Charlotte - South Corridor LRT	17134957.65	23232	28558262.75	0
2	Denver - Southeast Corridor - T-Rex	312572029.7	0	0	0
3	Denver Southwest Corridor	20842628	0	0	0
4	Los Angeles - East Side Extension	0	0	0	22704
5	Los Angeles - Long Beach Blue Line	26655743			32766
6	Minneapolis - Hiawatha Corridor	117305809	0	0	0
7	New Jersey - Hudson-Bergen MOS-2	728229000	47.69416188	1098000	0
8	New Jersey - Newark Rail Link	98695000	624.6909236	1345000	0
9	Southern New Jersey LRT	90622478	0	0	7920
10	Norfolk - Light Rail Transit	28674685	7920	10643	0
11	Phoenix - Central Phoenix/East Valley	7107610	0	0	0
12	Pittsburgh Light Rail Stage I	18674638	0	0	19694
13	Pittsburgh Light Rail Stage II	10239448	0	0	978
14	Pittsburgh - North Shore LRT Connector	0	0	0	0
15	Portland Interstate MAX	26532000	14203.2	33665000	1056
16	Portland MAX Segment I	0	52219.2	32112000	23232
17	Portland - South Corridor/Portland Mall	8899074	3585	891109	0
18	Portland/Westside/Hillsboro MAX	139993000	0	0	1848
19	Sacramento Folsom Corridor	37088543	0	0	7392
20	Sacramento South Corridor	30007683	0	0	0
21	Sacramento Stage I	32954922	0	0	0
22	Salt Lake City - Mid Jordan LRT	4350755	0	0	0
23	Salt Lake North South Corridor	50066532	0	0	13992
24	San Diego Mission Valley East	0	0	0	0
25	Santa Clara VTA North Corridor	44381582	0	0	0
26	Santa Clara VTA Tasman West	0	37525	98416648	0
27	St Louis St Clair Cnty Extension	99066435	0	0	0

	Variable #:	9	10	11	12
Project #	Project Name	Guideway_03_$	Guideway_04_Q	Guideway_04_$	Guideway_05_Q
1	Charlotte - South Corridor LRT	0	4752	39981567.85	9504
2	Denver - Southeast Corridor - T-Rex	0	0	0	0
3	Denver Southwest Corridor	0	858	1469462	5094
4	Los Angeles - East Side Extension	96847630	1108.8	6436405	0
5	Los Angeles - Long Beach Blue Line	24067146	10785	31155410	8459
6	Minneapolis - Hiawatha Corridor	0	0	0	0
7	New Jersey - Hudson-Bergen MOS-2	0	0	0	0
8	New Jersey - Newark Rail Link	0	0	0	0
9	Southern New Jersey LRT	9310008	0	0	0
10	Norfolk - Light Rail Transit	0	3168	25570861	0
11	Phoenix - Central Phoenix/East Valley	0	3060	22967395	0
12	Pittsburgh Light Rail Stage I	16901866	5012.28	4915157	0
13	Pittsburgh Light Rail Stage II	485500	3032	16640660	6204
14	Pittsburgh - North Shore LRT Connector	0	2956.8	26477000	0
15	Portland Interstate MAX	2503000	3907.2	30276000	0
16	Portland MAX Segment I	21846000	4012.8	9560000	23654
17	Portland - South Corridor/Portland Mall	0	3200	17004153	686
18	Portland/Westside/Hillsboro MAX	13362000	0	0	0
19	Sacramento Folsom Corridor	4269713	750	2973366	0
20	Sacramento South Corridor	0	650	4771138	0
21	Sacramento Stage I	0	5016	1752000	0
22	Salt Lake City - Mid Jordan LRT	0	0	11492206	0
23	Salt Lake North South Corridor	26713965	265	5083810	0
24	San Diego Mission Valley East	0	12896	41416985	5253
25	Santa Clara VTA North Corridor	0	0	0	0
26	Santa Clara VTA Tasman West	0	0	0	0
27	St Louis St Clair Cnty Extension	0	0	0	0

	Variable #:	13	14	15	16
Project #	Project Name	Guideway_05_$	Guideway_06_Q	Guideway_06_$	Guideway_07_Q
1	Charlotte - South Corridor LRT	28558262.75	0	0	0
2	Denver - Southeast Corridor - T-Rex	0	0	0	0
3	Denver Southwest Corridor	4216906	0	0	0
4	Los Angeles - East Side Extension	0	0	0	8976
5	Los Angeles - Long Beach Blue Line	5810205	3296	21647205	
6	Minneapolis - Hiawatha Corridor	0	17000	117748272	0
7	New Jersey - Hudson-Bergen MOS-2	0	0	0	0
8	New Jersey - Newark Rail Link	0	0	0	0
9	Southern New Jersey LRT	0	0	0	0
10	Norfolk - Light Rail Transit	123201	0	0	0
11	Phoenix - Central Phoenix/East Valley	0	0	0	0
12	Pittsburgh Light Rail Stage I	0	0	0	10720.85
13	Pittsburgh Light Rail Stage II	21117016	0	0	0
14	Pittsburgh - North Shore LRT Connector	0	1003.2	51301000	2217.6
15	Portland Interstate MAX	0	0	0	0
16	Portland MAX Segment I	19006000	0	0	0
17	Portland - South Corridor/Portland Mall	2012000	0	0	
18	Portland/Westside/Hillsboro MAX	0	153.12	685000	15470.4
19	Sacramento Folsom Corridor	0	0	0	0
20	Sacramento South Corridor	0	0	0	0
21	Sacramento Stage I	0	0	0	0
22	Salt Lake City - Mid Jordan LRT	0	0	0	0
23	Salt Lake North South Corridor	0	0	0	0
24	San Diego Mission Valley East	18557693	3279	42308590	0
25	Santa Clara VTA North Corridor	0	0	0	0
26	Santa Clara VTA Tasman West	0	0	0	0
27	St Louis St Clair Cnty Extension	0	0	0	0

	Variable #:	17	18	19	20
Project #	Project Name	Guideway_07_$	Guideway_08_Q	Guideway_08_$	Guideway_09_Q
1	Charlotte - South Corridor LRT	0	0	0	0
2	Denver - Southeast Corridor - T-Rex	0	0	0	0
3	Denver Southwest Corridor	0	1586	4126913	1456
4	Los Angeles - East Side Extension	234440128	0	0	63360
5	Los Angeles - Long Beach Blue Line		490	2183688	26324
6	Minneapolis - Hiawatha Corridor	0	0	0	34000
7	New Jersey - Hudson-Bergen MOS-2	0	0	0	0
8	New Jersey - Newark Rail Link	0	0	0	0
9	Southern New Jersey LRT	0	0	0	0
10	Norfolk - Light Rail Transit	0	2112		21380.47035
11	Phoenix - Central Phoenix/East Valley	0	104016	601213	3858
12	Pittsburgh Light Rail Stage I	145243891	11838	6084050	21468
13	Pittsburgh Light Rail Stage II	0	10961	32216144	4462
14	Pittsburgh - North Shore LRT Connector	60274000	158.4	4780000	12672
15	Portland Interstate MAX	0	0	0	7920
16	Portland MAX Segment I	0	270	1227000	0
17	Portland - South Corridor/Portland Mall	15082267		1950092	5470
18	Portland/Westside/Hillsboro MAX	180327000	0	0	0
19	Sacramento Folsom Corridor	0	0	0	1500
20	Sacramento South Corridor	0	0	0	1300
21	Sacramento Stage I	0	0	0	10032
22	Salt Lake City - Mid Jordan LRT	0	54907	2145991	0
23	Salt Lake North South Corridor	0	0	0	0
24	San Diego Mission Valley East	0	7541	28004633	47432
25	Santa Clara VTA North Corridor	0	0	0	0
26	Santa Clara VTA Tasman West	0	1875	7500000	74250
27	St Louis St Clair Cnty Extension	0	0	0	0

	Variable #:	21	22	23	24
Project #	Project Name	Guideway_09_$	Guideway_10_Q	Guideway_10_$	Guideway_11_Q
1	Charlotte - South Corridor LRT	0	0	0	0
2	Denver - Southeast Corridor - T-Rex	0	0	0	0
3	Denver Southwest Corridor	327331	0	0	91090
4	Los Angeles - East Side Extension	26056743	0	0	0
5	Los Angeles - Long Beach Blue Line	4981257	65532	8588412	146710
6	Minneapolis - Hiawatha Corridor	7117340	11500	5027216	76996
7	New Jersey - Hudson-Bergen MOS-2	0	0	0	0
8	New Jersey - Newark Rail Link	0	0	0	0
9	Southern New Jersey LRT	0	15840	3322296	279915
10	Norfolk - Light Rail Transit	7430559	39239.38298	9221255	17524.14667
11	Phoenix - Central Phoenix/East Valley	1715370	190768.8	104898426	19261
12	Pittsburgh Light Rail Stage I	4610360	39746	2184509	103180
13	Pittsburgh Light Rail Stage II	1798100	335	55000	49501
14	Pittsburgh - North Shore LRT Connector	9676000	0	0	0
15	Portland Interstate MAX	3587000	32430	16917000	20466
16	Portland MAX Segment I	0	23164	8892000	137194
17	Portland - South Corridor/Portland Mall	6903114	19702	8883990	64822
18	Portland/Westside/Hillsboro MAX	0	0	0	187218.24
19	Sacramento Folsom Corridor	420887	14784	4523656	78978
20	Sacramento South Corridor	675367	0	0	65228
21	Sacramento Stage I	248000	0	0	144342
22	Salt Lake City - Mid Jordan LRT	0	0	0	111936
23	Salt Lake North South Corridor	0	27984	2281905	130416
24	San Diego Mission Valley East	10302953	0	0	10506
25	Santa Clara VTA North Corridor	0	68404	6402209	96100
26	Santa Clara VTA Tasman West	5568750	0	0	0
27	St Louis St Clair Cnty Extension	0	0	0	183744

	Variable #:	25	26	27	28	29
Project #	Project Name	Guideway_11_$	Guideway_12_Q	Guideway_12_$	Station_01_Q	Station_01_$
1	Charlotte - South Corridor LRT	0	0	0	11	43763870
2	Denver - Southeast Corridor - T-Rex	0	0	0	13	35289766.61
3	Denver Southwest Corridor	8716761	92546	600000	5	19000000
4	Los Angeles - East Side Extension	0	0	0	6	19097869
5	Los Angeles - Long Beach Blue Line	10407580	238566	13222458	18	18932742
6	Minneapolis - Hiawatha Corridor	33658742	0	0	16	20761794
7	New Jersey - Hudson-Bergen MOS-2	0	0	0	0	0
8	New Jersey - Newark Rail Link	0	0	0	0	0
9	Southern New Jersey LRT	27529384	0	0	20	30607596
10	Norfolk - Light Rail Transit	3942933		2976829	10	179749
11	Phoenix - Central Phoenix/East Valley	3594836	58	10326536	32	59751520
12	Pittsburgh Light Rail Stage I	3804559	0	0	9	15733846
13	Pittsburgh Light Rail Stage II	8047204	54298	3534769	11	14740981
14	Pittsburgh - North Shore LRT Connector	0	3	3818000	0	0
15	Portland Interstate MAX	5040000	60816	4592000	10	8228000
16	Portland MAX Segment I	11481000	160358	2549000	30	11835000
17	Portland - South Corridor/Portland Mall	8559550	89994	32195	22	15932758
18	Portland/Westside/Hillsboro MAX	25646000	187218.24	544000	19	16736000
19	Sacramento Folsom Corridor	11766793	95262	2560901	10	19908378
20	Sacramento South Corridor	12698621	65228	813074	6	22454518
21	Sacramento Stage I	10011078	154374	1712400	28	10270000
22	Salt Lake City - Mid Jordan LRT	22783650	34	5850297	10	13495221
23	Salt Lake North South Corridor	8012158	0	0	16	14889489
24	San Diego Mission Valley East	4962830	57938	15621495	0	0
25	Santa Clara VTA North Corridor	8032209	164504	1605000	22	4914000
26	Santa Clara VTA Tasman West	0	74250	1138432	11	9780605
27	St Louis St Clair Cnty Extension	24674337	0	0	8	10098758

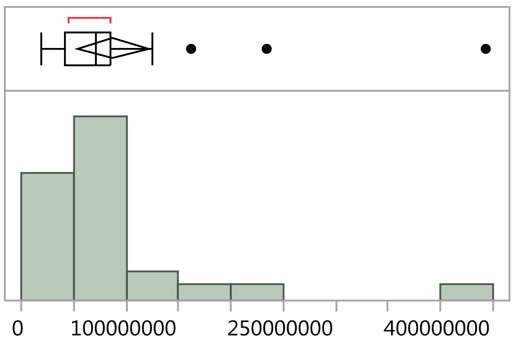
	Variable #:	30	31	32	33	34	35
Project #	Project Name	Station_02_Q	Station_02_$	Station_03_Q	Station_03_$	Station_06_Q	Station_06_$
1	Charlotte - South Corridor LRT	4	0	0	0		22846640
2	Denver - Southeast Corridor - T-Rex	0	0	0	0	700	12889245
3	Denver Southwest Corridor	0	0	0	0	0	0
4	Los Angeles - East Side Extension	0	0	2	78218510	0	0
5	Los Angeles - Long Beach Blue Line	3	8786682	1	27684300	0	0
6	Minneapolis - Hiawatha Corridor	0	0	1	18677389	0	0
7	New Jersey - Hudson-Bergen MOS-2	0	0	0	0	0	0
8	New Jersey - Newark Rail Link	0	0	0	0	0	0
9	Southern New Jersey LRT	0	0	0	0	0	0
10	Norfolk - Light Rail Transit		65620	0	0	0	0
11	Phoenix - Central Phoenix/East Valley	0	0	0	0	0	0
12	Pittsburgh Light Rail Stage I	0	0	3	18548933	0	0
13	Pittsburgh Light Rail Stage II					1538	22950000
14	Pittsburgh - North Shore LRT Connector	1	7417000	2	104541000	0	0
15	Portland Interstate MAX	0	0	0	0	600	1756000
16	Portland MAX Segment I	0	0	0	0	0	0
17	Portland - South Corridor/Portland Mall	0	0	0	0	750	16548246
18	Portland/Westside/Hillsboro MAX	0	0	1	44796000	880	9983000
19	Sacramento Folsom Corridor	0	0	0	0	0	0
20	Sacramento South Corridor	0	0	0	0	0	0
21	Sacramento Stage I	0	0	0	0	0	0
22	Salt Lake City - Mid Jordan LRT	0	0	0	0	0	0
23	Salt Lake North South Corridor	0	0	0	0	0	0
24	San Diego Mission Valley East	3	14919772	1	52240030	0	0
25	Santa Clara VTA North Corridor	0	0	0	0	0	0
26	Santa Clara VTA Tasman West	0	0	0	0	0	0
27	St Louis St Clair Cnty Extension	0	0	0	0	0	0

	Variable #:	36	37	38	39	40
Project #	Project Name	Support_Fac_$	Sitework_$	Utilities_$	Environmental_$	Accessways_$
1	Charlotte - South Corridor LRT	27366724	2807779	2176138	0	0
2	Denver - Southeast Corridor - T-Rex	3121471.88	37877960	7575383	0	36387398.52
3	Denver Southwest Corridor	400000	0	800000	600000	0
4	Los Angeles - East Side Extension	1230964	2640536	73998597	12934282	321974
5	Los Angeles - Long Beach Blue Line	44204740	13056748	1.22E+08		25468038
6	Minneapolis - Hiawatha Corridor	36647066	0	11243000	0	3632140
7	New Jersey - Hudson-Bergen MOS-2	0	29000	30072000	15083000	0
8	New Jersey - Newark Rail Link	0	0	14212000	244000	0
9	Southern New Jersey LRT	57927400	10763876	32700000	2673470	21654104
10	Norfolk - Light Rail Transit	12353577	27132636	15515744	2999574	10554092
11	Phoenix - Central Phoenix/East Valley	35573976	134599470	85617491	4808895	79650480
12	Pittsburgh Light Rail Stage I	37116525	747080	8974004	317890	0
13	Pittsburgh Light Rail Stage II	17249953	12572774	4063100	0	3984609
14	Pittsburgh - North Shore LRT Connector	0	34620000	23885000	2872000	7478000
15	Portland Interstate MAX	14306000	2198000	0	1000	31167000
16	Portland MAX Segment I	11672000	77672000	0	0	3912000
17	Portland - South Corridor/Portland Mall	7061320	109711510	30848859	1190879	31487206
18	Portland/Westside/Hillsboro MAX	19540000	13363000	13319000	3633000	10373332
19	Sacramento Folsom Corridor	8342257	2077081	11716000	4179328	0
20	Sacramento South Corridor	0	292465	7586004	0	0
21	Sacramento Stage I	3979000	6820334	5333092	0	0
22	Salt Lake City - Mid Jordan LRT	12528631	73595883	10078545	759269	38827386
23	Salt Lake North South Corridor	14579238	0	4141763	1272271	0
24	San Diego Mission Valley East	0	34676735	16012401	1075385	0
25	Santa Clara VTA North Corridor	21291136	8013000	5822000	2152000	0
26	Santa Clara VTA Tasman West	0	4237718	9074056	0	0
27	St Louis St Clair Cnty Extension	13203257	881457	4665032	3077702	0

	Variable #:	41	42	43	44	45	46
Project #	Project Name	Systems_$	ROW_Q	ROW_$	Vehicles_Q	Vehicles_$	Prof_Serv_$
1	Charlotte - South Corridor LRT	52429079	50688	3.7E+07	16	52384189	99080076
2	Denver - Southeast Corridor - T-Rex	163171396	100320	5.2E+07	34	85331149	128947559
3	Denver Southwest Corridor	11600000	44687	3.9E+07	18	32700000	33100000
4	Los Angeles - East Side Extension	94154817	31680	3.8E+07	10	22390323	168579703
5	Los Angeles - Long Beach Blue Line	115273245	119283	6E+07	54	78136129	877269855
6	Minneapolis - Hiawatha Corridor	73931596	61248	4.9E+07	26	74711000	70087000
7	New Jersey - Hudson-Bergen MOS-2	0	31680	3.8E+07	23	68420000	121858000
8	New Jersey - Newark Rail Link	0	46464	1.4E+07		13266000	64162000
9	Southern New Jersey LRT	49098264	147878	8.4E+07	20	73943650	147366035
10	Norfolk - Light Rail Transit	34523484	39072	1.6E+07	9	36066517	68054071
11	Phoenix - Central Phoenix/East Valley	144491764	105600	1.4E+08	50	116941301	308663033
12	Pittsburgh Light Rail Stage I	60031997	82198	2.2E+07	55	57399440	132804580
13	Pittsburgh Light Rail Stage II	47635624	28763	2950000	28	68400544	96918894
14	Pittsburgh - North Shore LRT Connector	42174000	6336	9422000	0	0	105188000
15	Portland Interstate MAX	36187000	30360	8396000	24	72618000	42116000
16	Portland MAX Segment I	21014000	103388	1.6E+07	26	25366000	47612000
17	Portland - South Corridor/Portland Mall	75578917	43507	2.1E+07	22	82334263	90231110
18	Portland/Westside/Hillsboro MAX	77431000	93609.1	6.8E+07	36	112853000	218533000
19	Sacramento Folsom Corridor	62594643	68070	8311529	14	35687292	51020907
20	Sacramento South Corridor	26878266	33264	2E+07	24	60908467	37198985
21	Sacramento Stage I	19514037	111936	1.7E+07	36	34600000	19034000
22	Salt Lake City - Mid Jordan LRT	72564520	55968	5.5E+07	28	107064677	48831474
23	Salt Lake North South Corridor	31084252	79200	3.5E+07	23	60250409	41955975
24	San Diego Mission Valley East	30670465	28969	3.4E+07	11	39218600	120284321
25	Santa Clara VTA North Corridor	33185990	82252	5.5E+07	50	55611000	87832000
26	Santa Clara VTA Tasman West	25019913	39400	3.7E+07	10	29418580	132769595
27	St Louis St Clair Cnty Extension	44554692	91872	2E+07	24	62094619	68025204

Appendix 4-C: Potential Model Variables - Data Distribution Analysis Plots

Cost_per_Mile



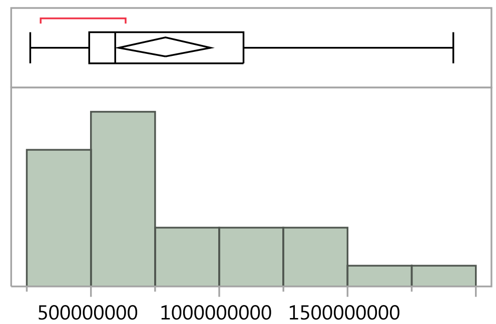
Quantiles

100.0%	maximum	443789164
99.5%		443789164
97.5%		443789164
90.0%		176247260.6
75.0%	quartile	84917408
50.0%	median	71187802
25.0%	quartile	40146602
10.0%		29518879.6
2.5%		18148233
0.5%		18148233
0.0%	minimum	18148233

Summary Statistics

Mean	87013944
Std Dev	84443837
Std Err Mean	16251224
Upper 95% Mean	120418813
Lower 95% Mean	53609074
N	27

Project_Total_Cost



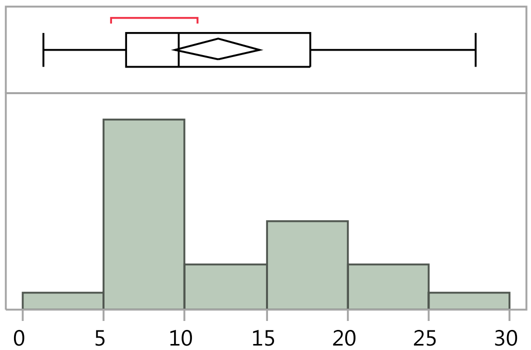
Quantiles

100.0%	maximum	1918271942
99.5%		1918271942
97.5%		1918271942
90.0%		1503345159.6
75.0%	quartile	1091296863
50.0%	median	593153369
25.0%	quartile	492520969
10.0%		326049517.8
2.5%		262474780
0.5%		262474780
0.0%	minimum	262474780

Summary Statistics

Mean	786495387
Std Dev	451981001
Std Err Mean	86983784
Upper 95% Mean	965293116
Lower 95% Mean	607697658
N	27

Alignment_Length



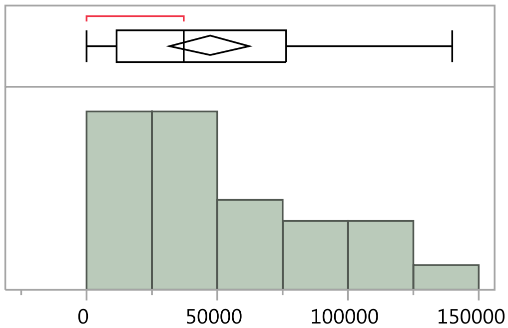
Quantiles

100.0%	maximum	28
99.5%		28
97.5%		28
90.0%		21.48
75.0%	quartile	17.7
50.0%	median	9.6
25.0%	quartile	6.3
10.0%		5.48
2.5%		1.2
0.5%		1.2
0.0%	minimum	1.2

Summary Statistics

Mean	11.992593
Std Dev	6.6429793
Std Err Mean	1.278442
Upper 95% Mean	14.620468
Lower 95% Mean	9.3647175
N	27

Guideway_01_Q



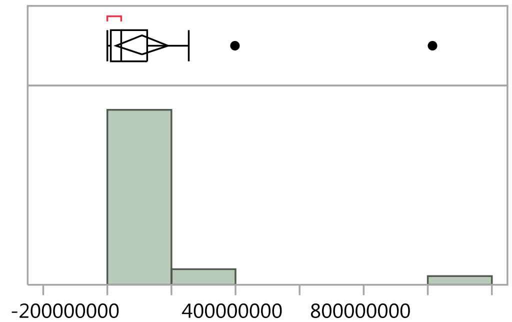
Quantiles

100.0%	maximum	139958
99.5%		139958
97.5%		139958
90.0%		104596.8
75.0%	quartile	76138
50.0%	median	37149
25.0%	quartile	11194
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	46902.667
Std Dev	39009.588
Std Err Mean	7507.3986
Upper 95% Mean	62334.346
Lower 95% Mean	31470.988
N	27

Guideway_01_$



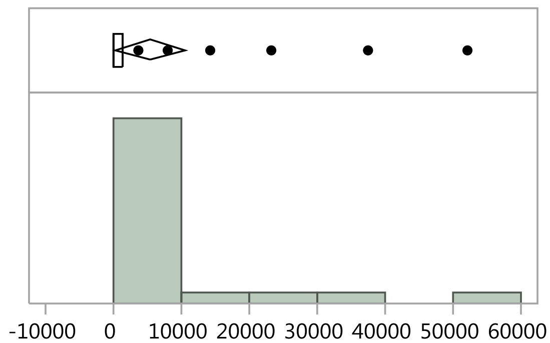
Quantiles

100.0%	maximum	1016660933
99.5%		1016660933
97.5%		1016660933
90.0%		280207168
75.0%	quartile	125622895
50.0%	median	43907758
25.0%	quartile	8124109
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	107272125
Std Dev	203251008
Std Err Mean	39115675
Upper 95% Mean	187675545
Lower 95% Mean	26868704
N	27

Guideway_02_Q



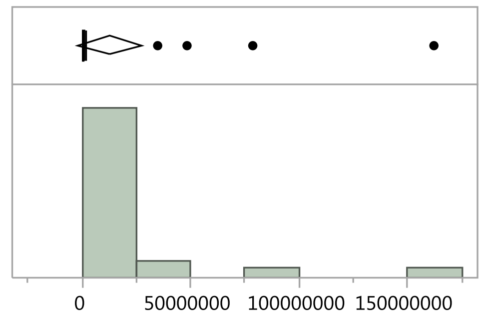
Quantiles

100.0%	maximum	52219
99.5%		52219
97.5%		52219
90.0%		27519.9
75.0%	quartile	1365
50.0%	median	0
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	5359.8846
Std Dev	12969.865
Std Err Mean	2543.5997
Upper 95% Mean	10598.526
Lower 95% Mean	121.24291
N	26

Guideway_02_$



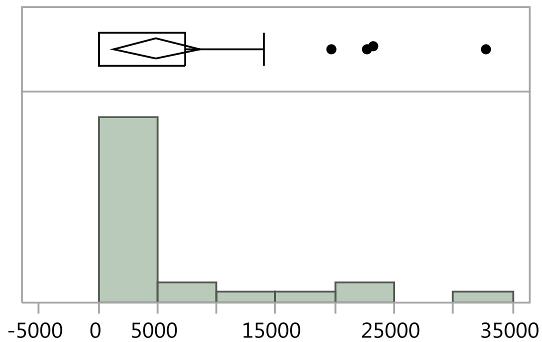
Quantiles

100.0%	maximum	162218328
99.5%		162218328
97.5%		162218328
90.0%		57406957.1
75.0%	quartile	1148408
50.0%	median	0
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	12623239
Std Dev	35723577
Std Err Mean	7005969.9
Upper 95% Mean	27052304
Lower 95% Mean	-1805826
N	26

Guideway_03_Q



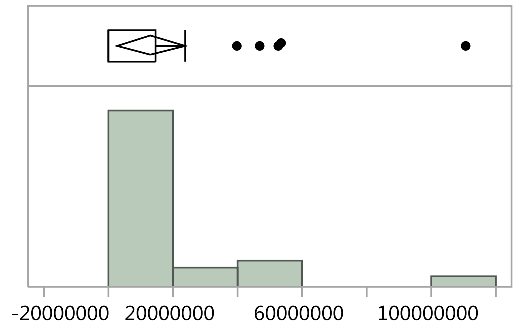
Quantiles

100.0%	maximum	32766
99.5%		32766
97.5%		32766
90.0%		22809.6
75.0%	quartile	7392
50.0%	median	0
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	4873.4074
Std Dev	9186.1936
Std Err Mean	1767.8838
Upper 95% Mean	8507.3446
Lower 95% Mean	1239.4702
N	27

Guideway_03_$



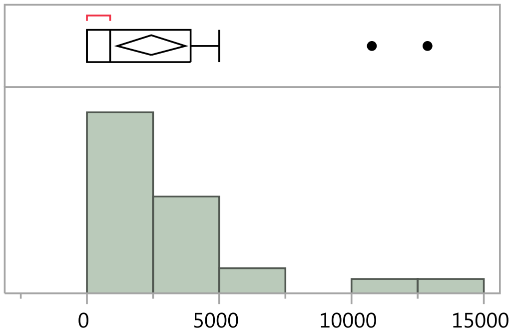
Quantiles

100.0%	maximum	110882739
99.5%		110882739
97.5%		110882739
90.0%		52800785.8
75.0%	quartile	14543361
50.0%	median	0
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	13043153
Std Dev	26266549
Std Err Mean	5054999.8
Upper 95% Mean	23433854
Lower 95% Mean	2652452.2
N	27

Guideway_04_Q



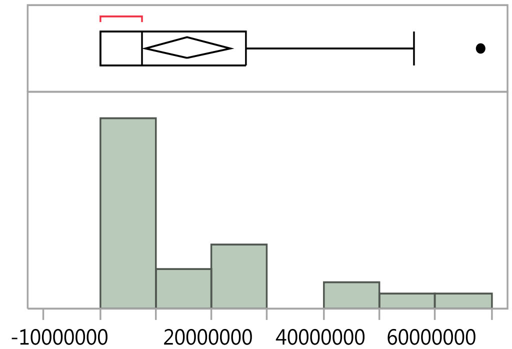
Quantiles

100.0%	maximum	12896
99.5%		12896
97.5%		12896
90.0%		6169.8
75.0%	quartile	3907
50.0%	median	858
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	2423.3333
Std Dev	3275.1572
Std Err Mean	630.30431
Upper 95% Mean	3718.9424
Lower 95% Mean	1127.7243
N	27

Guideway_04_$



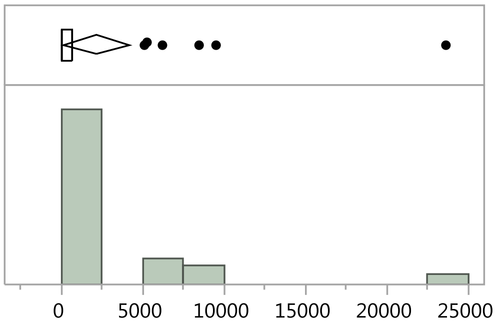
Quantiles

100.0%	maximum	68125616
99.5%		68125616
97.5%		68125616
90.0%		50200376
75.0%	quartile	26252089
50.0%	median	7453096
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	15687756
Std Dev	19291021
Std Err Mean	3712558.8
Upper 95% Mean	23319029
Lower 95% Mean	8056481.8
N	27

Guideway_05_Q



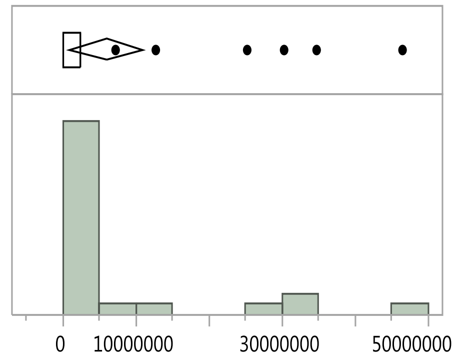
Quantiles

100.0%	maximum	23654
99.5%		23654
97.5%		23654
90.0%		8668
75.0%	quartile	686
50.0%	median	0
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	2179.7778
Std Dev	5135.1235
Std Err Mean	988.25498
Upper 95% Mean	4211.165
Lower 95% Mean	148.39057
N	27

Guideway_05_$



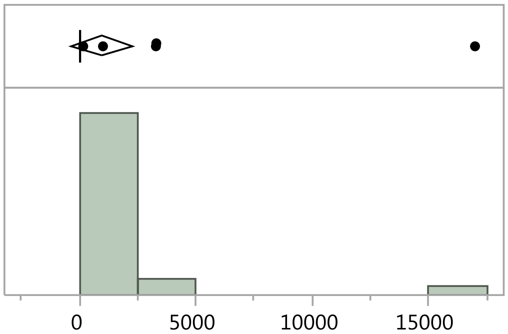
Quantiles

100.0%	maximum	46544385
99.5%		46544385
97.5%		46544385
90.0%		31193654.2
75.0%	quartile	2303578
50.0%	median	0
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	5895661.5
Std Dev	12725012
Std Err Mean	2448929.7
Upper 95% Mean	10929509
Lower 95% Mean	861814.31
N	27

Guideway_06_Q



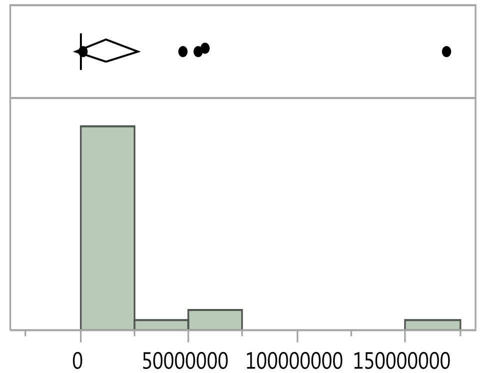
Quantiles

100.0%	maximum	17000
99.5%		17000
97.5%		17000
90.0%		3282.4
75.0%	quartile	0
50.0%	median	0
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	915.96296
Std Dev	3333.9293
Std Err Mean	641.61499
Upper 95% Mean	2234.8215
Lower 95% Mean	-402.8955
N	27

Guideway_06_$



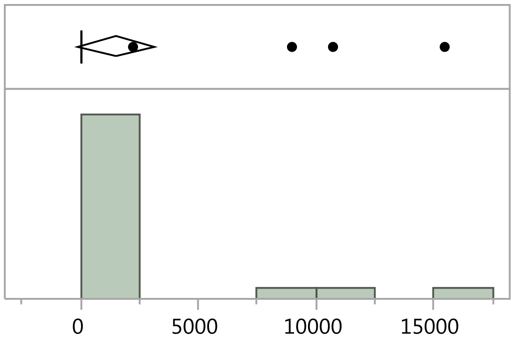
Quantiles

100.0%	maximum	168960986
99.5%		168960986
97.5%		168960986
90.0%		54996496
75.0%	quartile	0
50.0%	median	0
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	12201043
Std Dev	35646236
Std Err Mean	6860121.4
Upper 95% Mean	26302225
Lower 95% Mean	-1900138
N	27

Guideway_07_Q



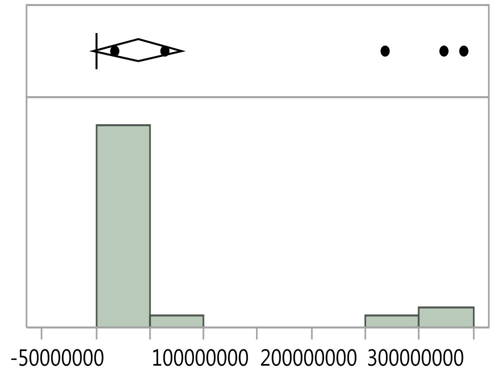
Quantiles

100.0%	maximum	15470
99.5%		15470
97.5%		15470
90.0%		9674
75.0%	quartile	0
50.0%	median	0
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	1495.4
Std Dev	3999.1838
Std Err Mean	799.83677
Upper 95% Mean	3146.182
Lower 95% Mean	-155.382
N	25

Guideway_07_$



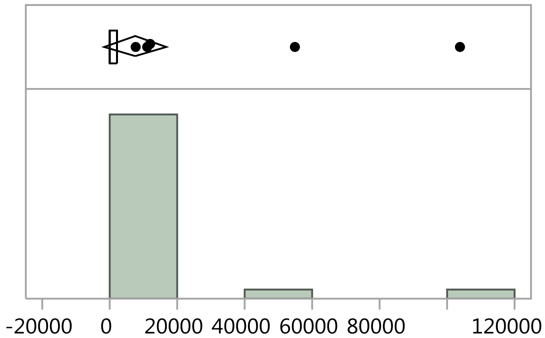
Quantiles

100.0%	maximum	341497039
99.5%		341497039
97.5%		341497039
90.0%		284806520.9
75.0%	quartile	0
50.0%	median	0
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	39003855
Std Dev	101558578
Std Err Mean	19917276
Upper 95% Mean	80024252
Lower 95% Mean	-2016542
N	26

Guideway_08_Q



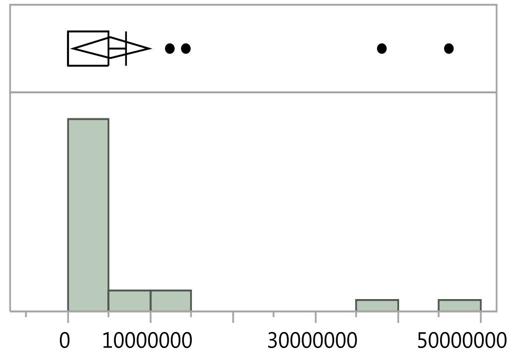
Quantiles

100.0%	maximum	104016
99.5%		104016
97.5%		104016
90.0%		24758.7
75.0%	quartile	1934.25
50.0%	median	0
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	7529
Std Dev	22528.429
Std Err Mean	4418.1885
Upper 95% Mean	16628.429
Lower 95% Mean	-1570.429
N	26

Guideway_08_$



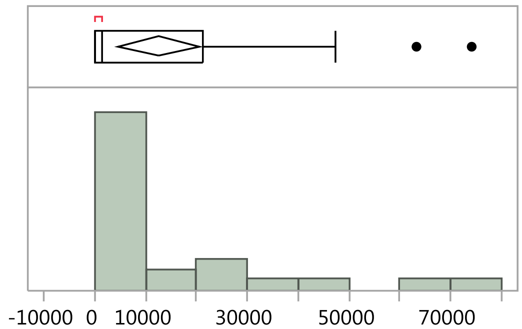
Quantiles

100.0%	maximum	46228037
99.5%		46228037
97.5%		46228037
90.0%		21440581.6
75.0%	quartile	4847435.25
50.0%	median	0
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	5234740.6
Std Dev	11577146
Std Err Mean	2270465.1
Upper 95% Mean	9910850.9
Lower 95% Mean	558630.26
N	26

Guideway_09_Q



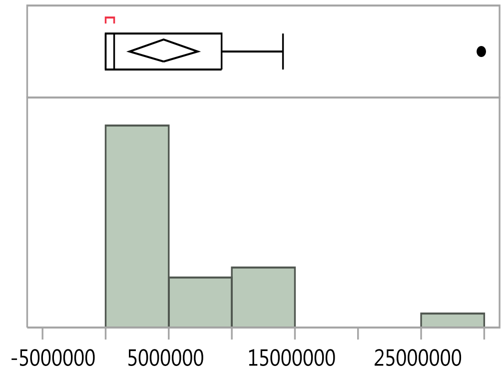
Quantiles

100.0%	maximum	74250
99.5%		74250
97.5%		74250
90.0%		50617.6
75.0%	quartile	21380
50.0%	median	1500
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	12477.185
Std Dev	20320.165
Std Err Mean	3910.6175
Upper 95% Mean	20515.575
Lower 95% Mean	4438.7957
N	27

Guideway_09_$



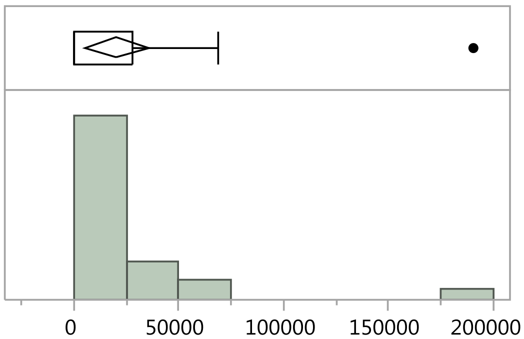
Quantiles

100.0%	maximum	29832873
99.5%		29832873
97.5%		29832873
90.0%		11516495.8
75.0%	quartile	9178867
50.0%	median	583097
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	4577665.7
Std Dev	6839317.6
Std Err Mean	1316227.3
Upper 95% Mean	7283209.6
Lower 95% Mean	1872121.8
N	27

Guideway_10_Q



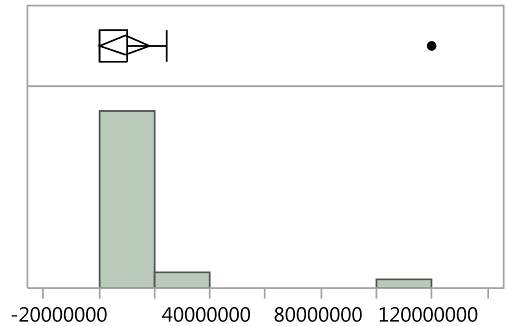
Quantiles

100.0%	maximum	190769
99.5%		190769
97.5%		190769
90.0%		66106.4
75.0%	quartile	27984
50.0%	median	0
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	20349.222
Std Dev	39548.77
Std Err Mean	7611.1644
Upper 95% Mean	35994.195
Lower 95% Mean	4704.2498
N	27

Guideway_10_$



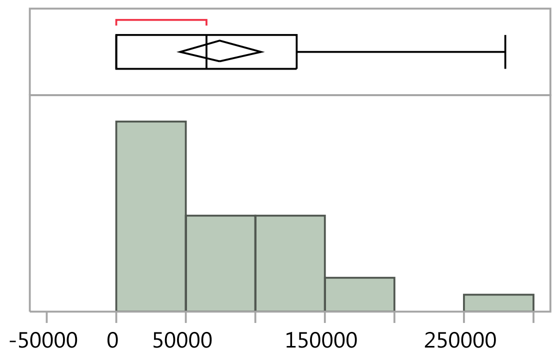
Quantiles

100.0%	maximum	119900529
99.5%		119900529
97.5%		119900529
90.0%		22275671.4
75.0%	quartile	9944680
50.0%	median	0
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	9173059.6
Std Dev	23294539
Std Err Mean	4483036.1
Upper 95% Mean	18388072
Lower 95% Mean	-41953.09
N	27

Guideway_11_Q



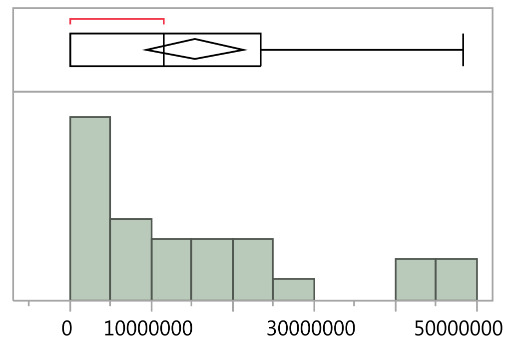
Quantiles

100.0%	maximum	279915
99.5%		279915
97.5%		279915
90.0%		184438.8
75.0%	quartile	130416
50.0%	median	65228
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	74634.333
Std Dev	73365.439
Std Err Mean	14119.185
Upper 95% Mean	103656.73
Lower 95% Mean	45611.932
N	27

Guideway_11_$



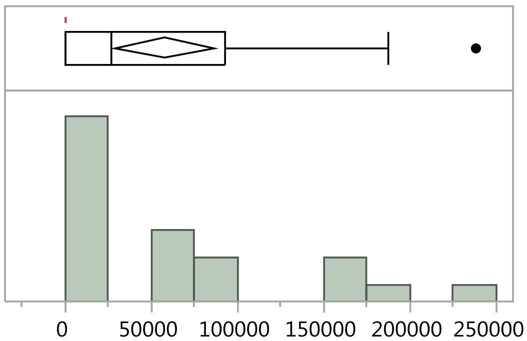
Quantiles

100.0%	maximum	48298070
99.5%		48298070
97.5%		48298070
90.0%		43592276.2
75.0%	quartile	23538019
50.0%	median	11547206
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	15302888
Std Dev	15136980
Std Err Mean	2913113.1
Upper 95% Mean	21290878
Lower 95% Mean	9314898.8
N	27

Guideway_12_Q



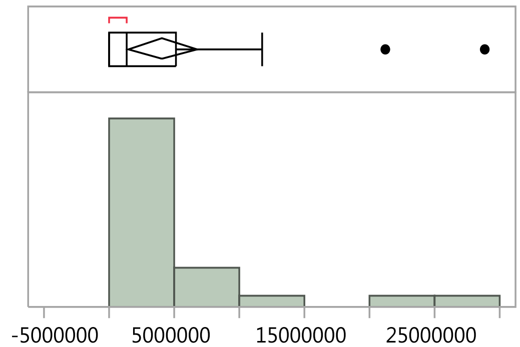
Quantiles

100.0%	maximum	238566
99.5%		238566
97.5%		238566
90.0%		171318.2
75.0%	quartile	93225
50.0%	median	27178
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	57517.192
Std Dev	71573.165
Std Err Mean	14036.653
Upper 95% Mean	86426.219
Lower 95% Mean	28608.165
N	26

Guideway_12_$



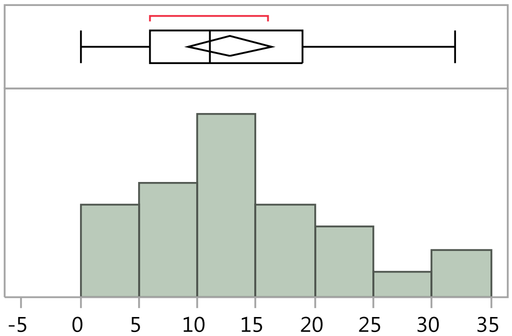
Quantiles

100.0%	maximum	28912734
99.5%		28912734
97.5%		28912734
90.0%		13692254
75.0%	quartile	5072160
50.0%	median	1270120
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	4070265.7
Std Dev	6804207.2
Std Err Mean	1309470.3
Upper 95% Mean	6761920.5
Lower 95% Mean	1378611
N	27

Station_01_Q



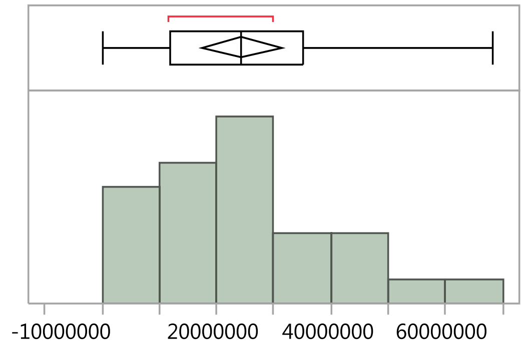
Quantiles

100.0%	maximum	32
99.5%		32
97.5%		32
90.0%		28.4
75.0%	quartile	19
50.0%	median	11
25.0%	quartile	6
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	12.703704
Std Dev	8.9649523
Std Err Mean	1.7253059
Upper 95% Mean	16.250121
Lower 95% Mean	9.1572867
N	27

Station_01_$



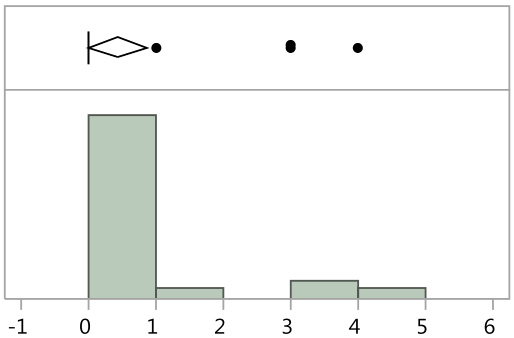
Quantiles

100.0%	maximum	68296915
99.5%		68296915
97.5%		68296915
90.0%		48904423
75.0%	quartile	35076679
50.0%	median	24146796
25.0%	quartile	11806636
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	24385100
Std Dev	17561056
Std Err Mean	3379626.9
Upper 95% Mean	31332023
Lower 95% Mean	17438178
N	27

Station_02_Q



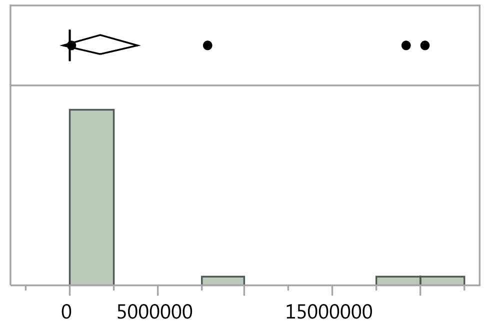
Quantiles

100.0%	maximum	4
99.5%		4
97.5%		4
90.0%		3
75.0%	quartile	0
50.0%	median	0
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	0.4230769
Std Dev	1.1017469
Std Err Mean	0.2160703
Upper 95% Mean	0.8680821
Lower 95% Mean	-0.021928
N	26

Station_02_$



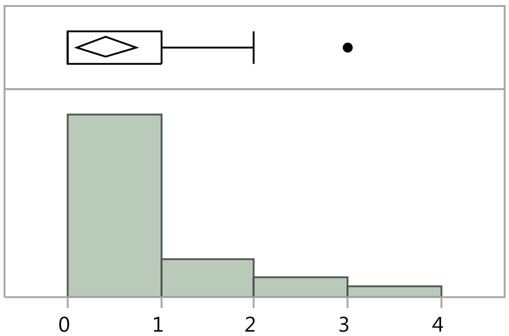
Quantiles

100.0%	maximum	20293252
99.5%		20293252
97.5%		20293252
90.0%		10129957.6
75.0%	quartile	0
50.0%	median	0
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	1756905.1
Std Dev	5404401.1
Std Err Mean	1040077.5
Upper 95% Mean	3894815
Lower 95% Mean	-381004.7
N	27

Station_03_Q



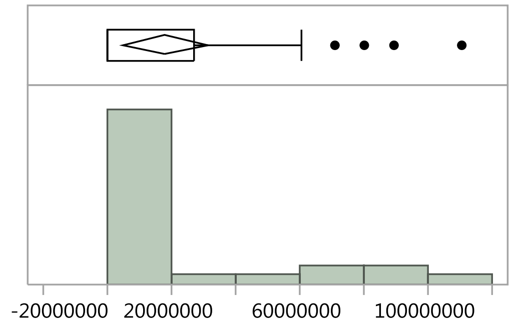
Quantiles

100.0%	maximum	3
99.5%		3
97.5%		3
90.0%		2
75.0%	quartile	1
50.0%	median	0
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	0.4074074
Std Dev	0.7970744
Std Err Mean	0.153397
Upper 95% Mean	0.7227195
Lower 95% Mean	0.0920953
N	27

Station_03_$



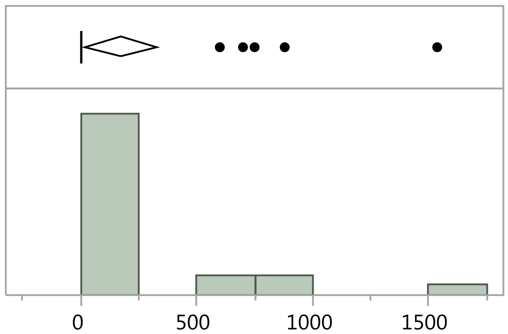
Quantiles

100.0%	maximum	110772632
99.5%		110772632
97.5%		110772632
90.0%		82111893
75.0%	quartile	26800818
50.0%	median	0
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	17873378
Std Dev	33649231
Std Err Mean	6475797.5
Upper 95% Mean	31184570
Lower 95% Mean	4562185.5
N	27

Station_06_Q



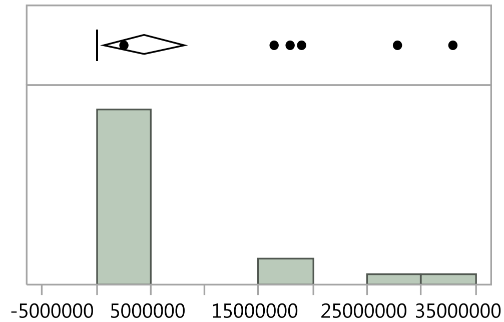
Quantiles

100.0%	maximum	1538
99.5%		1538
97.5%		1538
90.0%		789
75.0%	quartile	0
50.0%	median	0
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	171.84615
Std Dev	389.07765
Std Err Mean	76.304404
Upper 95% Mean	328.99802
Lower 95% Mean	14.694291
N	26

Station_06_$



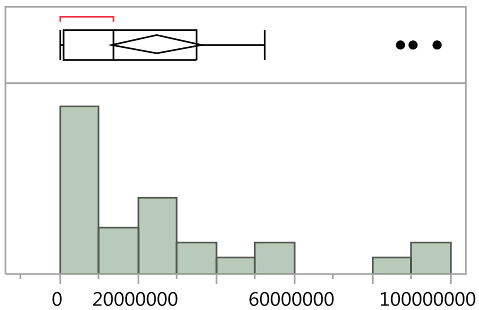
Quantiles

100.0%	maximum	32931733
99.5%		32931733
97.5%		32931733
90.0%		20719083.8
75.0%	quartile	0
50.0%	median	0
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	4314741.1
Std Dev	9424898.5
Std Err Mean	1813822.6
Upper 95% Mean	8043106.8
Lower 95% Mean	586375.44
N	27

Support_Fac_$



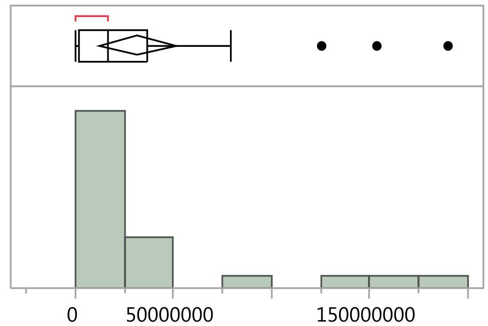
Quantiles

100.0%	maximum	96659781
99.5%		96659781
97.5%		96659781
90.0%		87912541.6
75.0%	quartile	35005630
50.0%	median	13511526
25.0%	quartile	682217
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	24774773
Std Dev	28738065
Std Err Mean	5530643.3
Upper 95% Mean	36143173
Lower 95% Mean	13406373
N	27

Sitework_$



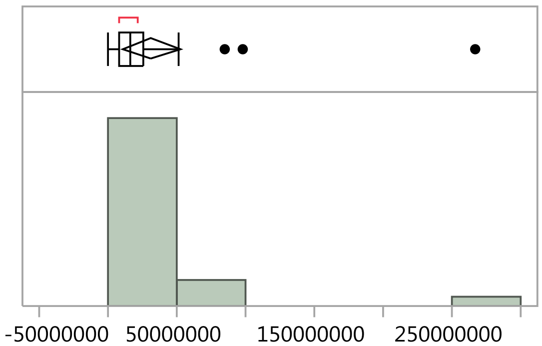
Quantiles

100.0%	maximum	190213378
99.5%		190213378
97.5%		190213378
90.0%		131258533.4
75.0%	quartile	36683679
50.0%	median	16035951
25.0%	quartile	1452889
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	31692591
Std Dev	49715139
Std Err Mean	9567683
Upper 95% Mean	51359245
Lower 95% Mean	12025936
N	27

Utilities_$



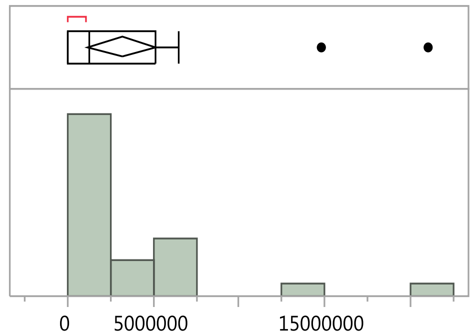
Quantiles

100.0%	maximum	267458981
99.5%		267458981
97.5%		267458981
90.0%		87350373
75.0%	quartile	25308772
50.0%	median	15935615
25.0%	quartile	7689285
10.0%		1091546.4
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	30952150
Std Dev	52757423
Std Err Mean	10153171
Upper 95% Mean	51822291
Lower 95% Mean	10082009
N	27

Environmental_$



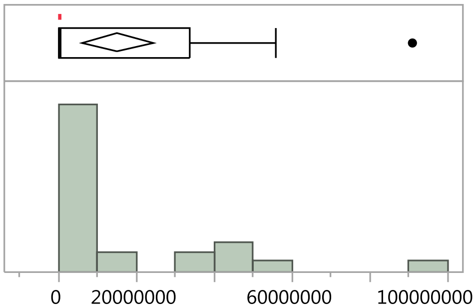
Quantiles

100.0%	maximum	21056971
99.5%		21056971
97.5%		21056971
90.0%		8998540.5
75.0%	quartile	5063062.5
50.0%	median	1193393
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	3157729.3
Std Dev	4932258.1
Std Err Mean	967295.39
Upper 95% Mean	5149911.4
Lower 95% Mean	1165547.1
N	26

Accessways_$



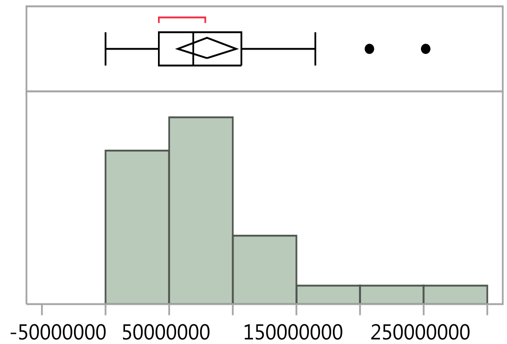
Quantiles

100.0%	maximum	91041735
99.5%		91041735
97.5%		91041735
90.0%		48190132.4
75.0%	quartile	33826335
50.0%	median	368634
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	15121742
Std Dev	23409242
Std Err Mean	4505110.8
Upper 95% Mean	24382130
Lower 95% Mean	5861354.6
N	27

Systems_$



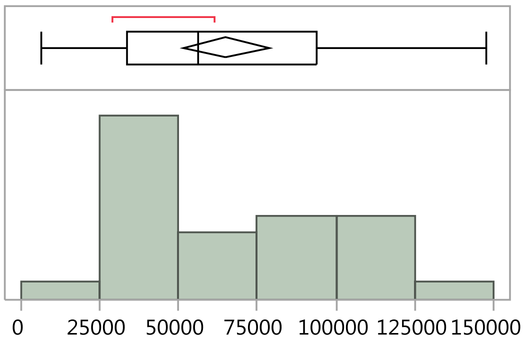
Quantiles

100.0%	maximum	252060903
99.5%		252060903
97.5%		252060903
90.0%		173663263.2
75.0%	quartile	106086953
50.0%	median	68353971
25.0%	quartile	41987109
10.0%		15827428
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	79975624
Std Dev	58723790
Std Err Mean	11301399
Upper 95% Mean	103205982
Lower 95% Mean	56745267
N	27

ROW_Q



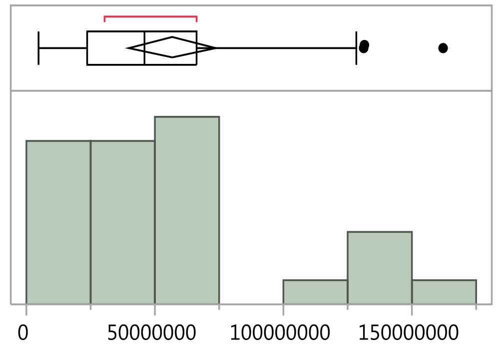
Quantiles

100.0%	maximum	147878
99.5%		147878
97.5%		147878
90.0%		113405.4
75.0%	quartile	93609
50.0%	median	55968
25.0%	quartile	33264
10.0%		28927.8
2.5%		6336
0.5%		6336
0.0%	minimum	6336

Summary Statistics

Mean	65099.704
Std Dev	35114.812
Std Err Mean	6757.8487
Upper 95% Mean	78990.661
Lower 95% Mean	51208.747
N	27

ROW_$



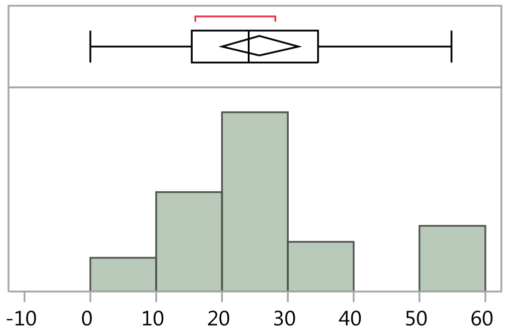
Quantiles

100.0%	maximum	162409720
99.5%		162409720
97.5%		162409720
90.0%		131444978
75.0%	quartile	66279194
50.0%	median	45893425
25.0%	quartile	23521035
10.0%		11040724
2.5%		4233055
0.5%		4233055
0.0%	minimum	4233055

Summary Statistics

Mean	56900448
Std Dev	42800804
Std Err Mean	8237018.5
Upper 95% Mean	73831882
Lower 95% Mean	39969014
N	27

Vehicles_Q



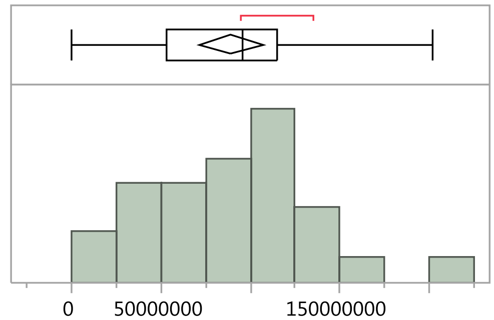
Quantiles

100.0%	maximum	55
99.5%		55
97.5%		55
90.0%		51.2
75.0%	quartile	34.5
50.0%	median	24
25.0%	quartile	15.5
10.0%		9.7
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	25.807692
Std Dev	14.338812
Std Err Mean	2.8120725
Upper 95% Mean	31.599264
Lower 95% Mean	20.016121
N	26

Vehicles_$



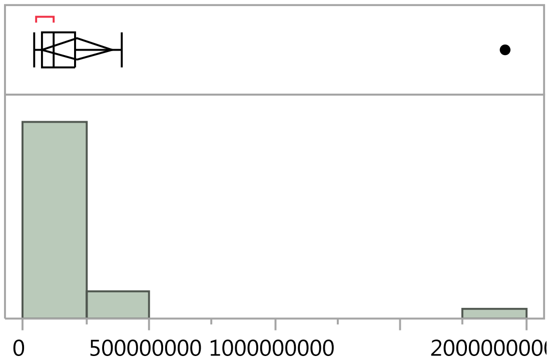
Quantiles

100.0%	maximum	202174532
99.5%		202174532
97.5%		202174532
90.0%		142137016.6
75.0%	quartile	115464102
50.0%	median	95519323
25.0%	quartile	53343505
10.0%		23885191.4
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	89232232
Std Dev	46024003
Std Err Mean	8857323.5
Upper 95% Mean	107438721
Lower 95% Mean	71025743
N	27

Prof_Serv_$



Quantiles

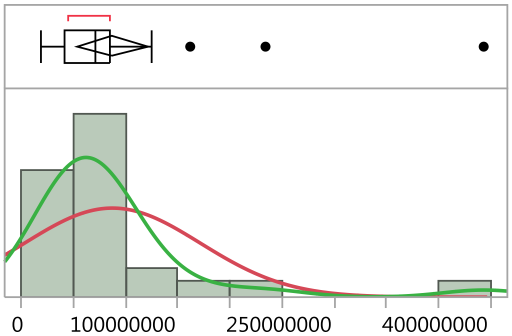
100.0%	maximum	1918271942
99.5%		1918271942
97.5%		1918271942
90.0%		360545057
75.0%	quartile	206510358
50.0%	median	116598508
25.0%	quartile	73393041
10.0%		55695225.4
2.5%		44752689
0.5%		44752689
0.0%	minimum	44752689

Summary Statistics

Mean	210941504
Std Dev	353421818
Std Err Mean	68016061
Upper 95% Mean	350750519
Lower 95% Mean	71132489
N	27

Appendix 4-D: Potential Model Variables – Fit Curve Analysis Plots

Cost_per_Mile



Normal(8.7e+7,8.44e+7)

Smooth Curve

Nonparametric Density

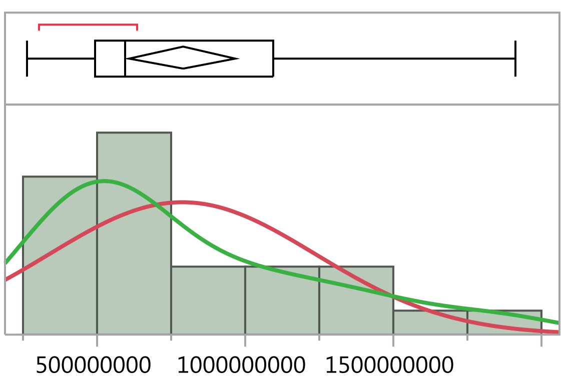
Kernel Std
39313138

Fitted Normal - Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	87013944	53609074	120418813
Dispersion	σ	84443837	66500908	115724503

-2log(Likelihood) = 1061.20893045543

Project_Total_Cost



Normal(7.86e+8,4.52e+8)

Smooth Curve

Nonparametric Density

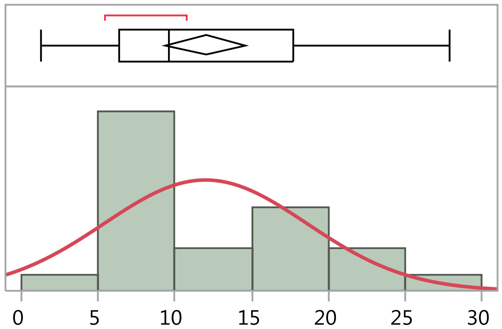
Kernel Std
2.104e+8

Fitted Normal - Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	786495387	607697658	965293116
Dispersion	σ	451981001	355942459	619409054

-2log(Likelihood) = 1151.79681885157

Alignment_Length



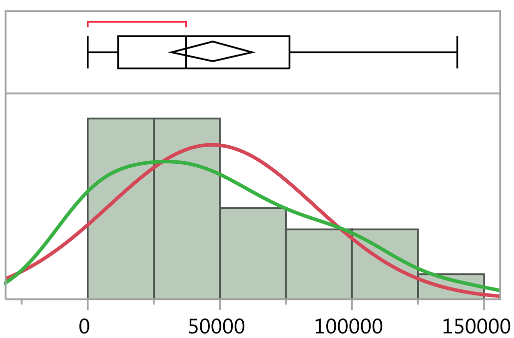
Normal(11.9926,6.64298)

Fitted Normal Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	11.992593	9.3647175	14.620468
Dispersion	σ	6.6429793	5.2314552	9.1037488

-2log(Likelihood) = 177.874950517434

Guideway_01_Q



Normal(46902.7,39009.6)

Smooth Curve

Nonparametric Density

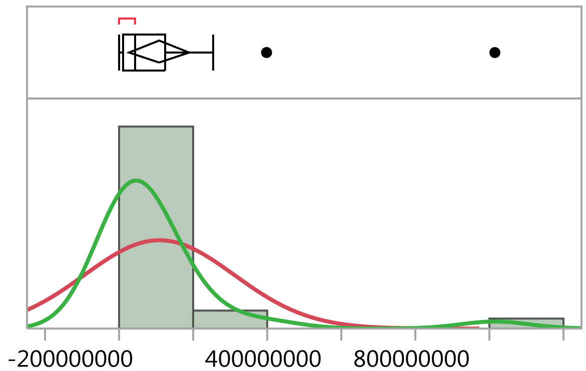
Kernel Std
18161.06

Fitted Normal - Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	46902.667	31470.988	62334.346
Dispersion	σ	39009.588	30720.691	53459.972

-2log(Likelihood) = 646.487068360628

Guideway_01_$



Normal(1.07e+8,2.03e+8)

Smooth Curve

Nonparametric Density

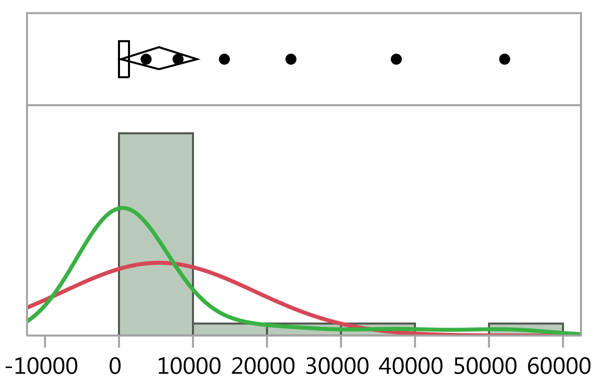
Kernel Std
94624253

Fitted Normal - Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	107272125	26868704	187675545
Dispersion	σ	203251008	160063506	278541607

-2log(Likelihood) = 1108.64010315818

Guideway_02_Q



Normal(5359.88,12969.9)

Smooth Curve

Nonparametric Density

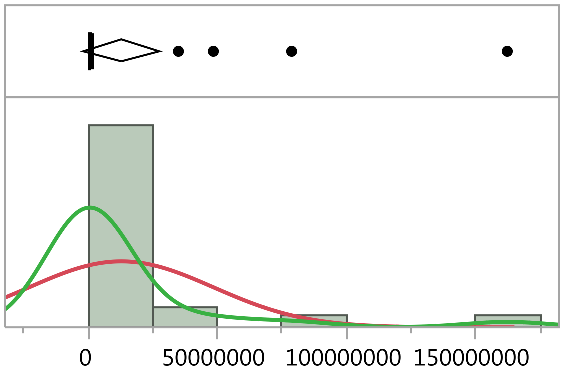
Kernel Std
6083.917

Fitted Normal - Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	5359.8846	121.24291	10598.526
Dispersion	σ	12969.865	10171.712	17903.715

-2log(Likelihood) = 565.244763532925

Guideway_02_$



Normal(1.26e+7,3.57e+7)

Smooth Curve **Nonparametric Density**

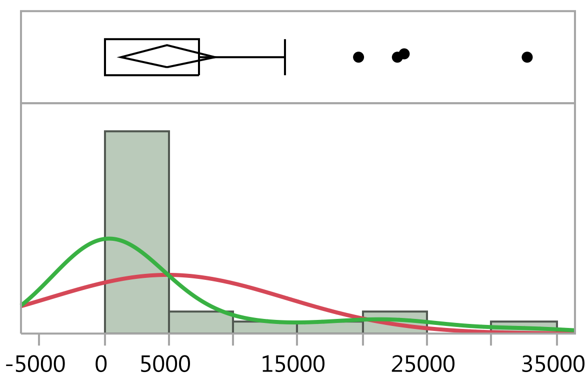
Kernel Std
16757251

Fitted Normal - Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	12623239	-1805826	27052304
Dispersion	σ	35723577	28016478	49313140

-2log(Likelihood) = 977.133519284353

Guideway_03_Q



Normal(4873.41,9186.19)

Smooth Curve **Nonparametric Density**

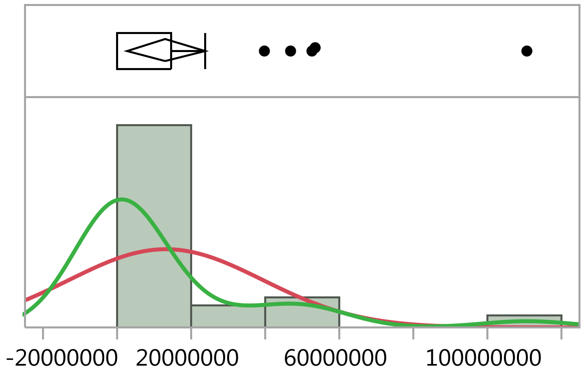
Kernel Std
4276.666

Fitted Normal - Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	4873.4074	1239.4702	8507.3446
Dispersion	σ	9186.1936	7234.2783	12589.05

-2log(Likelihood) = 568.397355764486

Guideway_03_$



Normal(1.3e+7,2.63e+7)

Smooth Curve

Nonparametric Density

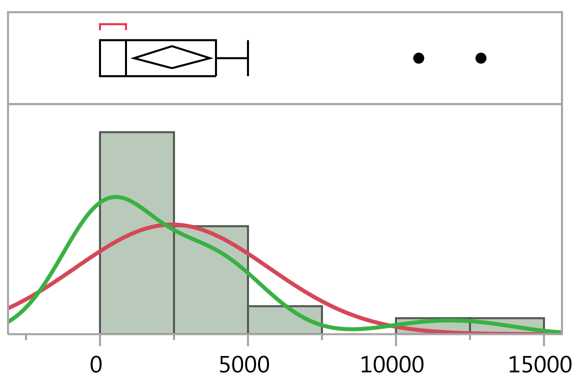
Kernel Std
12228489

Fitted Normal - Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	13043153	2652452.2	23433854
Dispersion	σ	26266549	20685339	35996510

-2log(Likelihood) = 998.148248263256

Guideway_04_Q



Normal(2423.33,3275.16)

Smooth Curve

Nonparametric Density

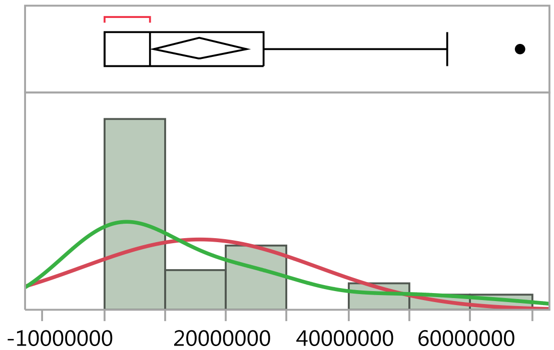
Kernel Std
1524.761

Fitted Normal - Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	2423.3333	1127.7243	3718.9424
Dispersion	σ	3275.1572	2579.2401	4488.379

-2log(Likelihood) = 512.705223520201

Guideway_04_$



Normal(1.57e+7,1.93e+7)

Smooth Curve

Nonparametric Density

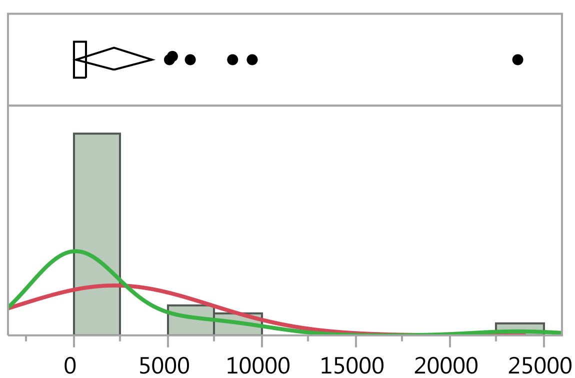
Kernel Std
8981006

Fitted Normal - Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	15687756	8056481.8	23319029
Dispersion	σ	19291021	15191996	26437025

-2log(Likelihood) = 981.480798252232

Guideway_05_Q



Normal(2179.78,5135.12)

Smooth Curve

Nonparametric Density

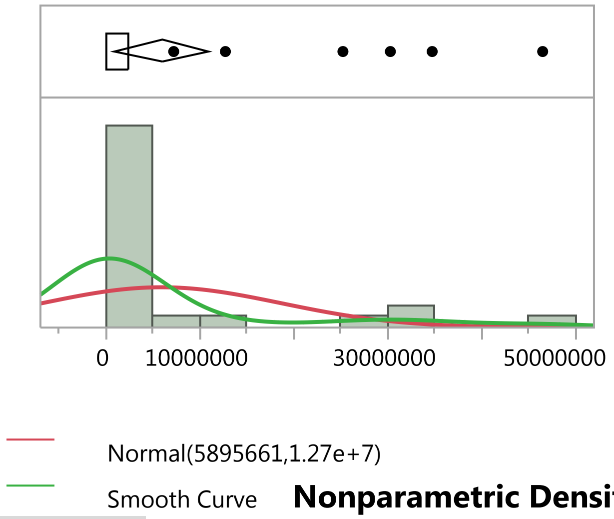
Kernel Std
2390.676

Fitted Normal - Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	2179.7778	148.39057	4211.165
Dispersion	σ	5135.1235	4043.9941	7037.3356

-2log(Likelihood) = 536.991076170395

Guideway_05_$



Nonparametric Density

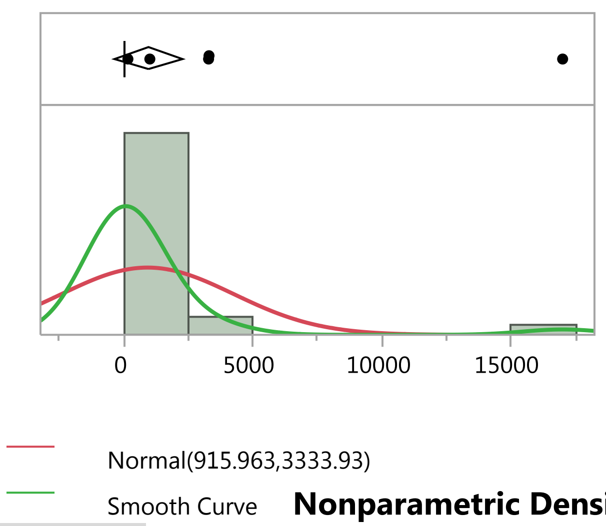
Kernel Std
5924176

Fitted Normal - Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	5895661.5	861814.31	10929509
Dispersion	σ	12725012	10021156	17438759

-2log(Likelihood) = 959.013005094479

Guideway_06_Q



Nonparametric Density

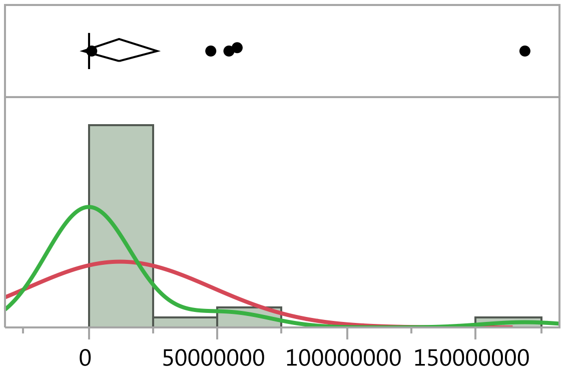
Kernel Std
1552.123

Fitted Normal - Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	915.96296	-402.8955	2234.8215
Dispersion	σ	3333.9293	2625.524	4568.9221

-2log(Likelihood) = 513.665650710385

Guideway_06_$



Normal(1.22e+7,3.56e+7)

Smooth Curve

Nonparametric Density

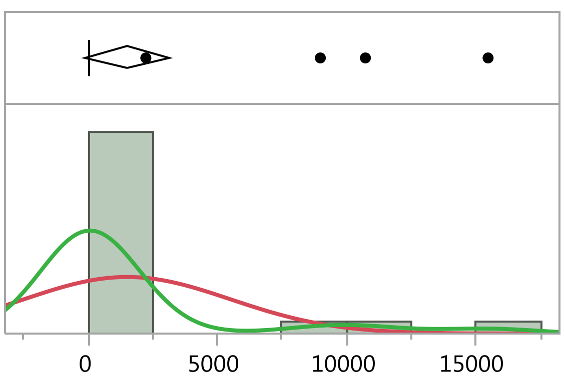
Kernel Std
16595236

Fitted Normal - Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	12201043	-1900138	26302225
Dispersion	σ	35646236	28071996	48850729

-2log(Likelihood) = 1014.63700344636

Guideway_07_Q



Normal(1495.4,3999.18)

Smooth Curve

Nonparametric Density

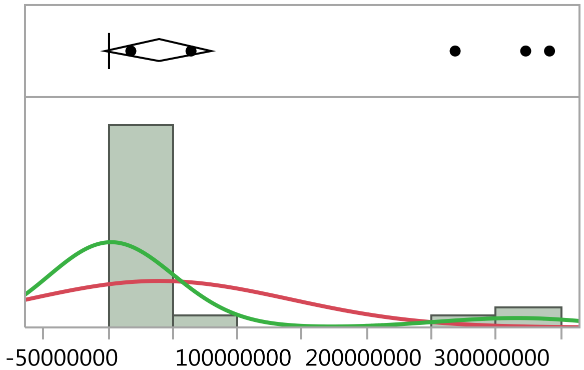
Kernel Std
1890.714

Fitted Normal - Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	1495.4	-155.382	3146.182
Dispersion	σ	3999.1838	3122.6762	5563.4733

-2log(Likelihood) = 484.639205541877

Guideway_07_$



Normal(3.9e+7,1.02e+8)

Smooth Curve

Nonparametric Density

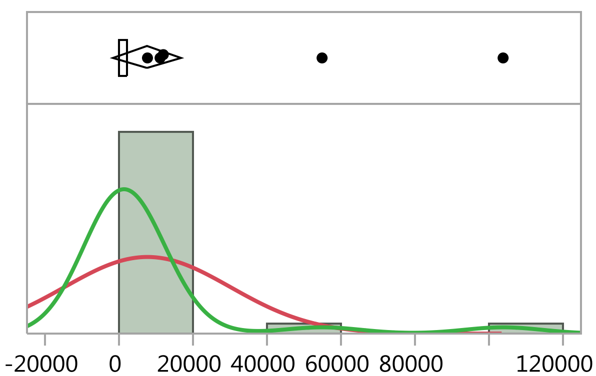
Kernel Std
47639199

Fitted Normal - Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	39003855	-2016542	80024252
Dispersion	σ	101558578	79648060	140192355

-2log(Likelihood) = 1031.46441180992

Guideway_08_Q



Normal(7529,22528.4)

Smooth Curve

Nonparametric Density

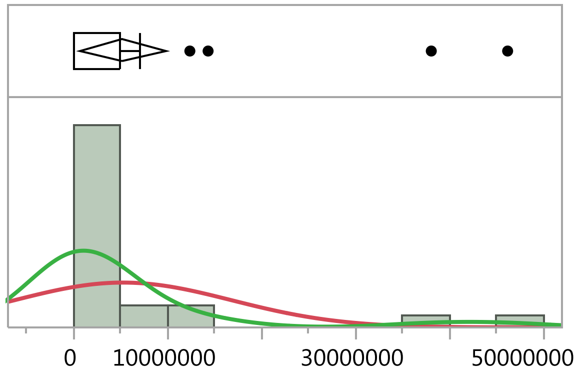
Kernel Std
10567.66

Fitted Normal - Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	7529	-1570.429	16628.429
Dispersion	σ	22528.429	17668.086	31098.442

-2log(Likelihood) = 593.956535794023

Guideway_08_$



Normal(5234741,1.16e+7)

Smooth Curve

Nonparametric Density

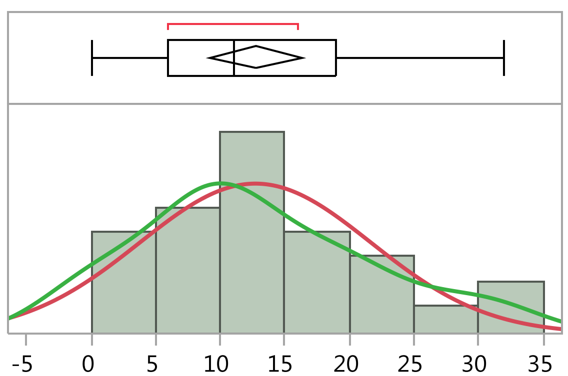
Kernel Std
5430619

Fitted Normal - Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	5234740.6	558630.26	9910850.9
Dispersion	σ	11577146	9079461.4	15981194

-2log(Likelihood) = 918.541066169125

Station_01_Q



Normal(12.7037,8.96495)

Smooth Curve

Nonparametric Density

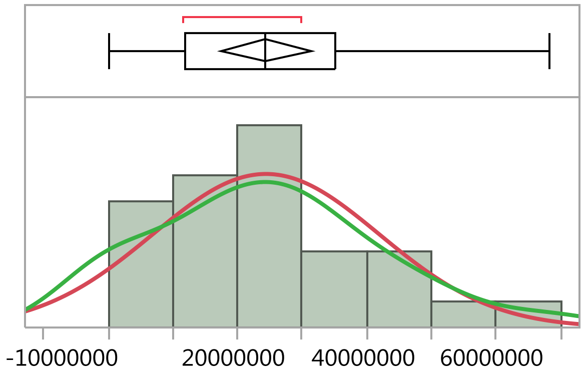
Kernel Std
4.173666

Fitted Normal - Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	12.703704	9.1572867	16.250121
Dispersion	σ	8.9649523	7.0600471	12.285854

-2log(Likelihood) = 194.062111467367

Station_01_$



Normal(2.44e+7,1.76e+7)

Smooth Curve

Nonparametric Density

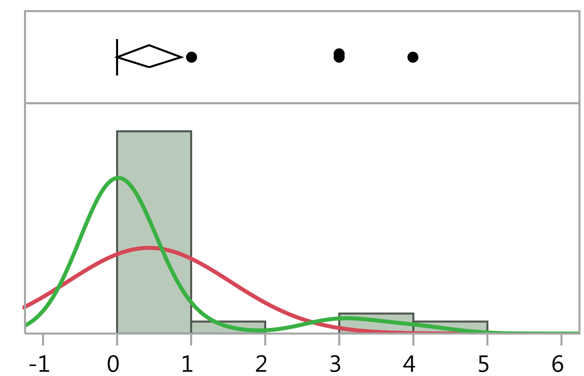
Kernel Std
8175614

Fitted Normal - Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	24385100	17438178	31332023
Dispersion	σ	17561056	13829620	24066227

-2log(Likelihood) = 976.407172870659

Station_02_Q



Normal(0.42308,1.10175)

Smooth Curve

Nonparametric Density

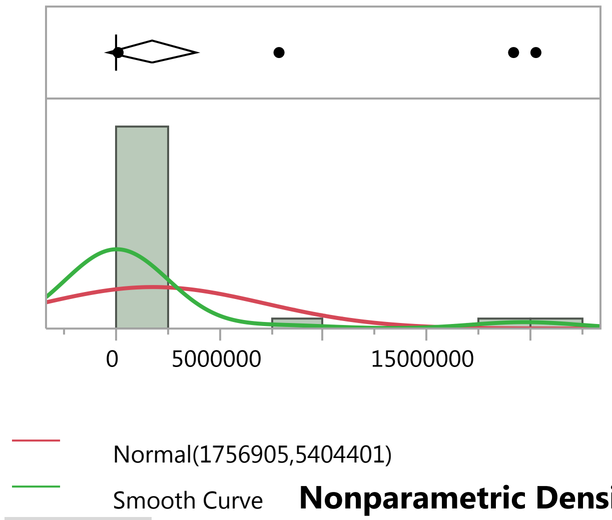
Kernel Std
0.516809

Fitted Normal - Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	0.4230769	-0.021928	0.8680821
Dispersion	σ	1.1017469	0.8640531	1.5208611

-2log(Likelihood) = 77.823446633504

Station_02_$



Nonparametric Density

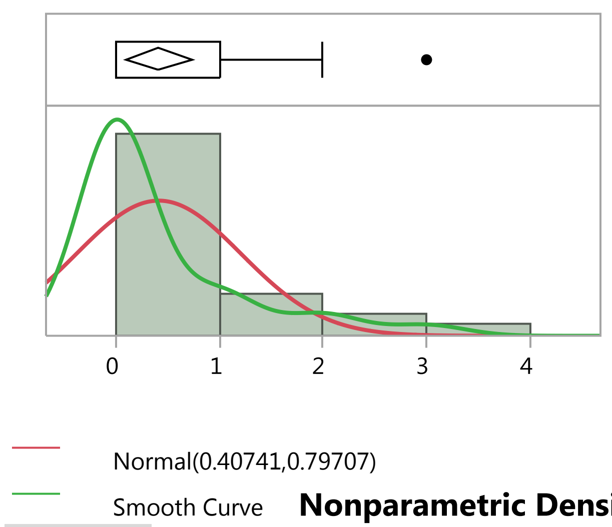
Kernel Std
2516039

Fitted Normal - Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	µ	1756905.1	-381004.7	3894815
Dispersion	σ	5404401.1	4256054.6	7406362.1

-2log(Likelihood) = 912.769786992199

Station_03_Q



Nonparametric Density

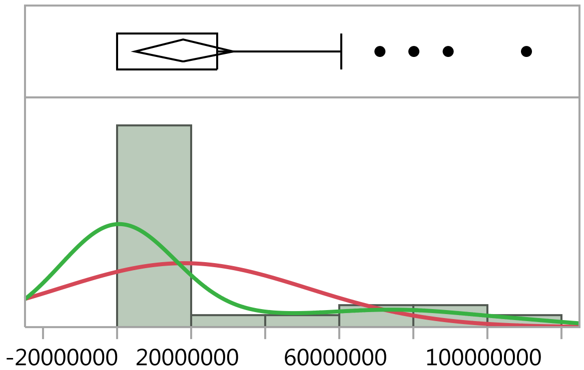
Kernel Std
0.371081

Fitted Normal - Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	µ	0.4074074	0.0920953	0.7227195
Dispersion	σ	0.7970744	0.6277092	1.092336

-2log(Likelihood) = 63.3750905888972

Station_03_$



Normal(1.79e+7,3.36e+7)

Smooth Curve

Nonparametric Density

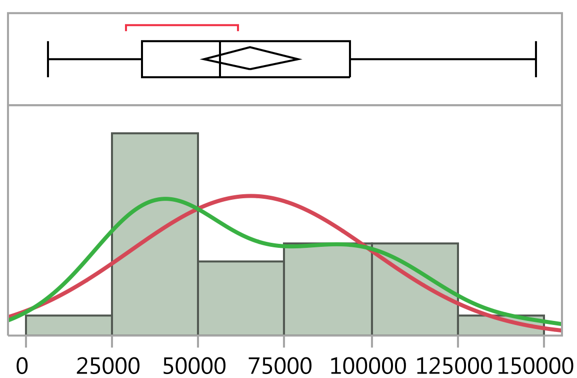
Kernel Std
15665523

Fitted Normal - Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	17873378	4562185.5	31184570
Dispersion	σ	33649231	26499322	46113970

-2log(Likelihood) = 1011.52372168121

ROW_Q



Normal(65099.7,35114.8)

Smooth Curve

Nonparametric Density

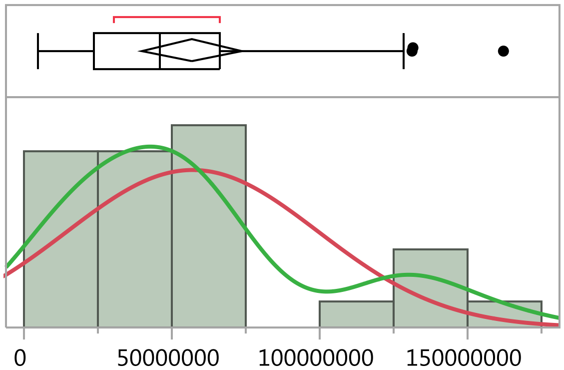
Kernel Std
16347.83

Fitted Normal - Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	65099.704	51208.747	78990.661
Dispersion	σ	35114.812	27653.491	48122.448

-2log(Likelihood) = 640.807109291312

ROW_$



Normal(5.69e+7,4.28e+7)

Smooth Curve

Nonparametric Density

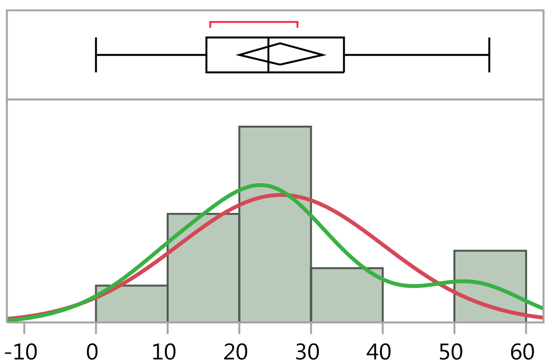
Kernel Std
19926071

Fitted Normal - Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	56900448	39969014	73831882
Dispersion	σ	42800804	33706336	58655575

-2log(Likelihood) = 1024.51432260779

Vehicles_Q



Normal(25.8077,14.3388)

Smooth Curve

Nonparametric Density

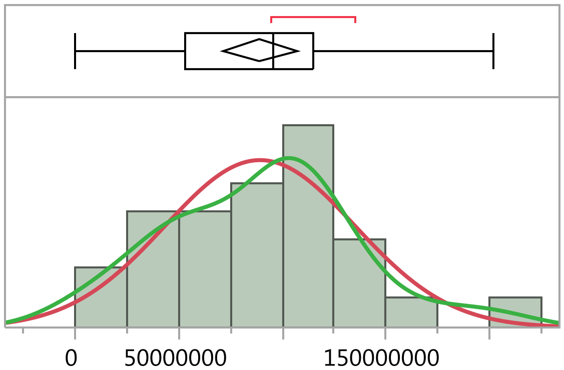
Kernel Std
6.726064

Fitted Normal - Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	25.807692	20.016121	31.599264
Dispersion	σ	14.338812	11.245319	19.793423

-2log(Likelihood) = 211.259244151551

Vehicles_$



Normal(8.92e+7,4.6e+7)

Smooth Curve

Nonparametric Density

Kernel Std
21426644

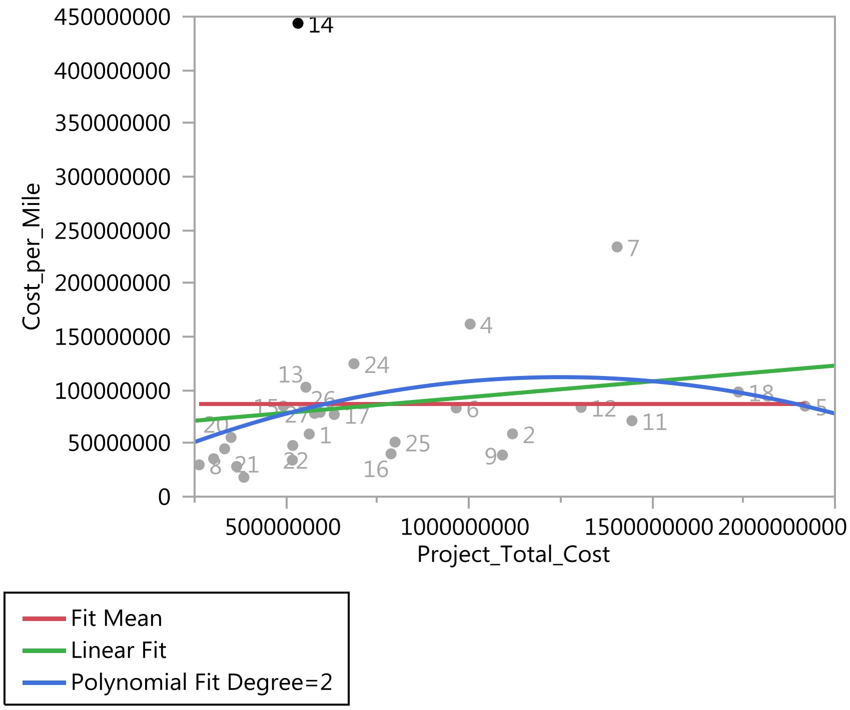
Fitted Normal - Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	89232232	71025743	107438721
Dispersion	σ	46024003	36244658	63072749

-2log(Likelihood) = 1028.43505653827

Appendix 4-E: Potential Model Variables – Linear vs Non-linear Analysis Plots

Bivariate Fit of Cost_per_Mile By Project_Total_Cost



Linear Fit

Cost_per_Mile = 63795932 + 0.0295208*Project_Total_Cost

Summary of Fit

RSquare	0.024967
RSquare Adj	-0.01403
Root Mean Square Error	85034339
Mean of Response	87013944
Observations (or Sum Wgts)	27

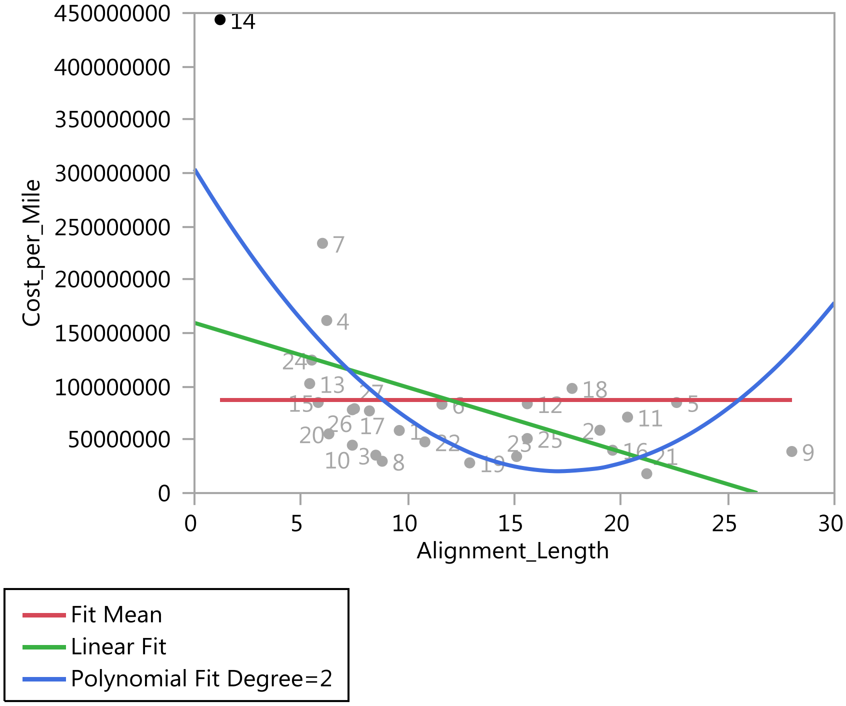
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	4.6288e+15	4.629e+15	0.6402
Error	25	1.8077e+17	7.231e+15	**Prob > F**
C. Total	26	1.854e+17		0.4312

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	63795932	33315401	1.91	0.0670
Project_Total_Cost	0.0295208	0.036897	0.80	0.4312

Bivariate Fit of Cost_per_Mile By Alignment_Length



Linear Fit

Cost_per_Mile = 159550529 - 6048449*Alignment_Length

Summary of Fit

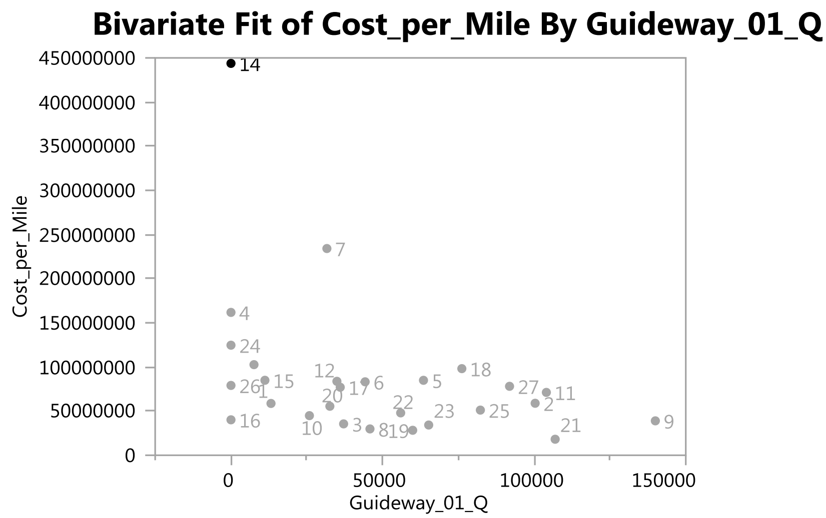
RSquare	0.226401
RSquare Adj	0.195457
Root Mean Square Error	75743023
Mean of Response	87013944
Observations (or Sum Wgts)	27

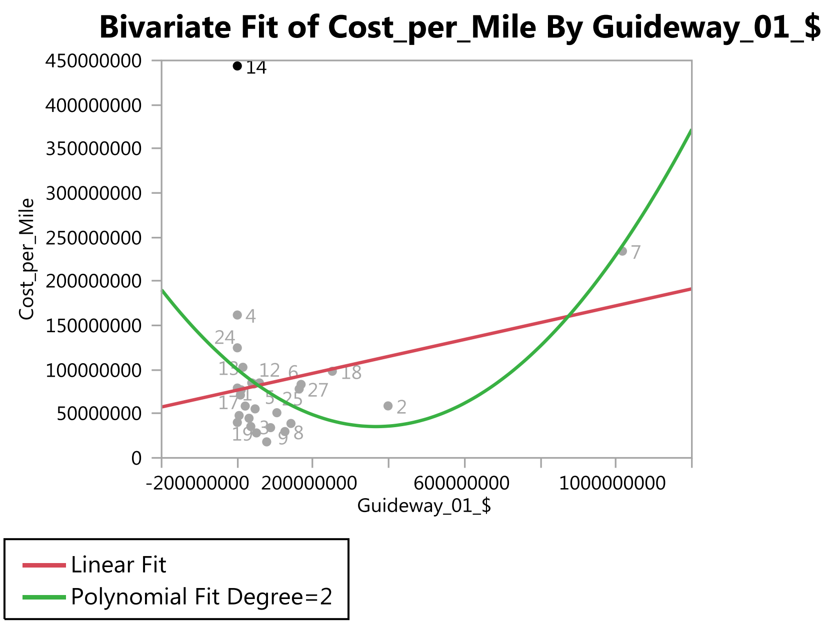
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	4.1975e+16	4.197e+16	7.3165
Error	25	1.4343e+17	5.737e+15	**Prob > F**
C. Total	26	1.854e+17		0.0121*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	159550529	30522447	5.23	<.0001*
Alignment_Length	-6048449	2236109	-2.70	0.0121*





Linear Fit

Cost_per_Mile = 76762578 + 0.0955641*Guideway_01_$

Summary of Fit

RSquare	0.052908
RSquare Adj	0.015024
Root Mean Square Error	83807092
Mean of Response	87013944
Observations (or Sum Wgts)	27

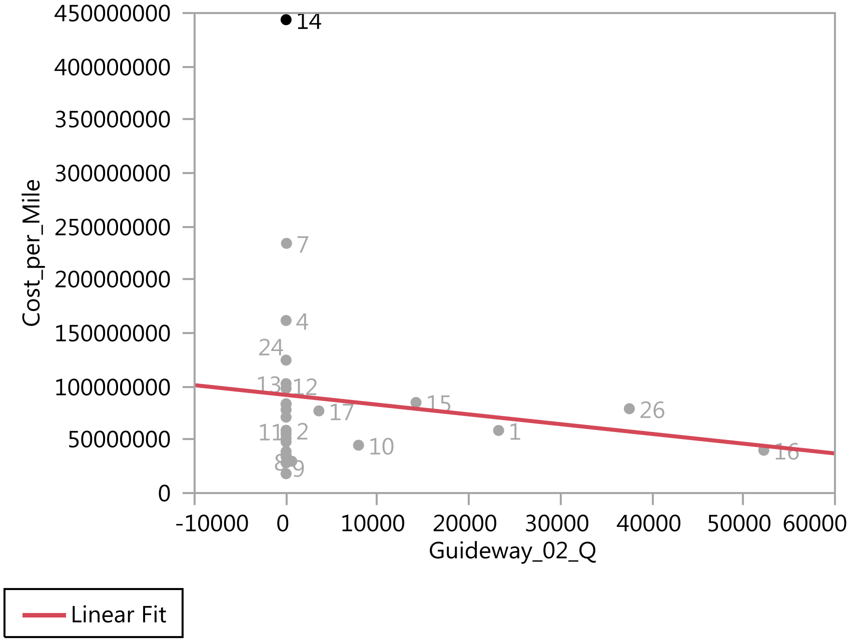
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	9.8091e+15	9.809e+15	1.3966
Error	25	1.7559e+17	7.024e+15	**Prob > F**
C. Total	26	1.854e+17		0.2484

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	76762578	18313456	4.19	0.0003*
Guideway_01_$	0.0955641	0.080865	1.18	0.2484

Bivariate Fit of Cost_per_Mile By Guideway_02_Q



Linear Fit

Cost_per_Mile = 91993577 - 913.73825*Guideway_02_Q

Summary of Fit

RSquare	0.018939
RSquare Adj	-0.02194
Root Mean Square Error	87054551
Mean of Response	87096046
Observations (or Sum Wgts)	26

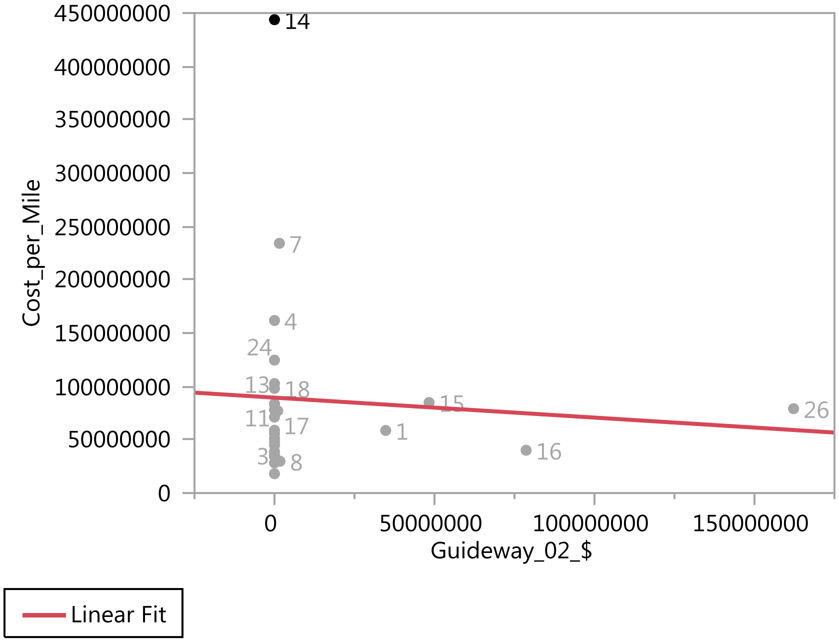
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	3.5112e+15	3.511e+15	0.4633
Error	24	1.8188e+17	7.578e+15	**Prob > F**
C. Total	25	1.854e+17		0.5026

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	91993577	18527038	4.97	<.0001*
Guideway_02_Q	-913.7382	1342.413	-0.68	0.5026

Bivariate Fit of Cost_per_Mile By Guideway_02_$



Linear Fit

Cost_per_Mile = 89457331 - 0.1870586*Guideway_02_$

Summary of Fit

RSquare	0.006022
RSquare Adj	-0.03539
Root Mean Square Error	87625792
Mean of Response	87096046
Observations (or Sum Wgts)	26

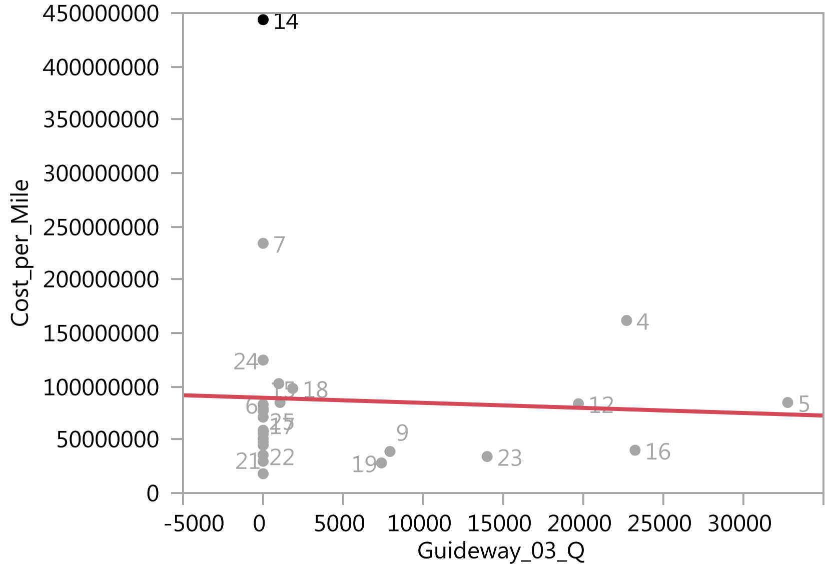
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1.1164e+15	1.116e+15	0.1454
Error	24	1.8428e+17	7.678e+15	**Prob > F**
C. Total	25	1.854e+17		0.7063

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	89457331	18266570	4.90	<.0001*
Guideway_02_$	-0.187059	0.490577	-0.38	0.7063

Bivariate Fit of Cost_per_Mile By Guideway_03_Q



Linear Fit

Cost_per_Mile = 89335303 - 476.3319*Guideway_03_Q

Summary of Fit

RSquare	0.002685
RSquare Adj	-0.03721
Root Mean Square Error	86000463
Mean of Response	87013944
Observations (or Sum Wgts)	27

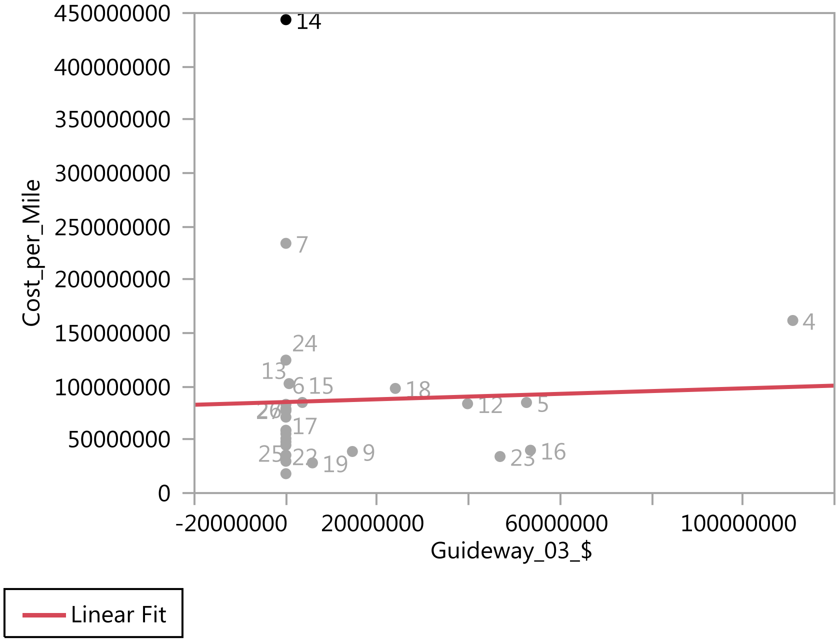
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	4.9781e+14	4.978e+14	0.0673
Error	25	1.849e+17	7.396e+15	**Prob > F**
C. Total	26	1.854e+17		0.7974

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	89335303	18814626	4.75	<.0001*
Guideway_03_Q	-476.3319	1836.025	-0.26	0.7974

Bivariate Fit of Cost_per_Mile By Guideway_03_$



Linear Fit

Cost_per_Mile = 85343133 + 0.1280987*Guideway_03_$

Summary of Fit

RSquare	0.001588
RSquare Adj	-0.03835
Root Mean Square Error	86047765
Mean of Response	87013944
Observations (or Sum Wgts)	27

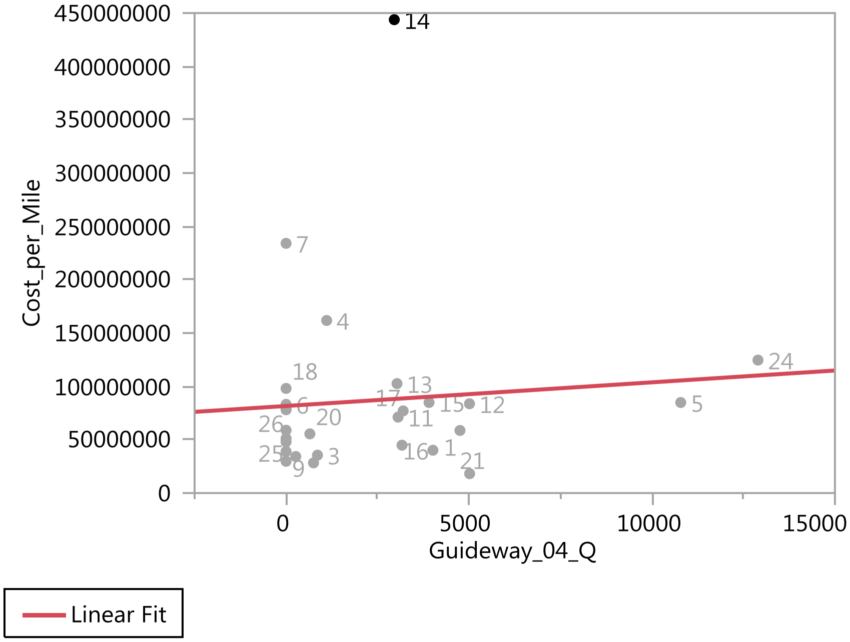
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	2.9435e+14	2.944e+14	0.0398
Error	25	1.8511e+17	7.404e+15	**Prob > F**
C. Total	26	1.854e+17		0.8436

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	85343133	18559390	4.60	0.0001*
Guideway_03_$	0.1280987	0.642466	0.20	0.8436

Bivariate Fit of Cost_per_Mile By Guideway_04_Q



Linear Fit

Cost_per_Mile = 81643229 + 2216.251*Guideway_04_Q

Summary of Fit

RSquare	0.007389
RSquare Adj	-0.03232
Root Mean Square Error	85797422
Mean of Response	87013944
Observations (or Sum Wgts)	27

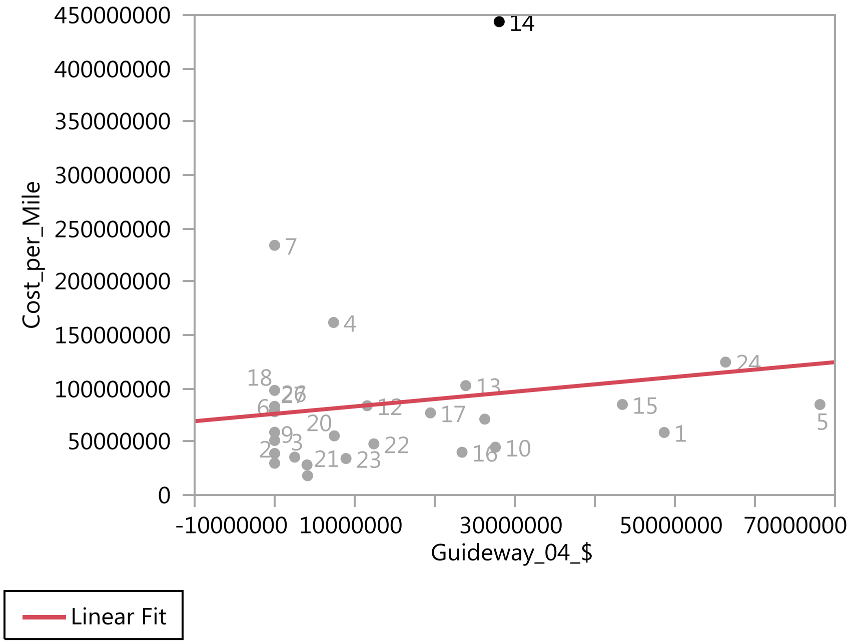
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1.3699e+15	1.37e+15	0.1861
Error	25	1.8403e+17	7.361e+15	**Prob > F**
C. Total	26	1.854e+17		0.6699

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	81643229	20679432	3.95	0.0006*
Guideway_04_Q	2216.251	5137.542	0.43	0.6699

Bivariate Fit of Cost_per_Mile By Guideway_04_$



Linear Fit

Cost_per_Mile = 76165948 + 0.6914944*Guideway_04_$

Summary of Fit

RSquare	0.024955
RSquare Adj	-0.01405
Root Mean Square Error	85034865
Mean of Response	87013944
Observations (or Sum Wgts)	27

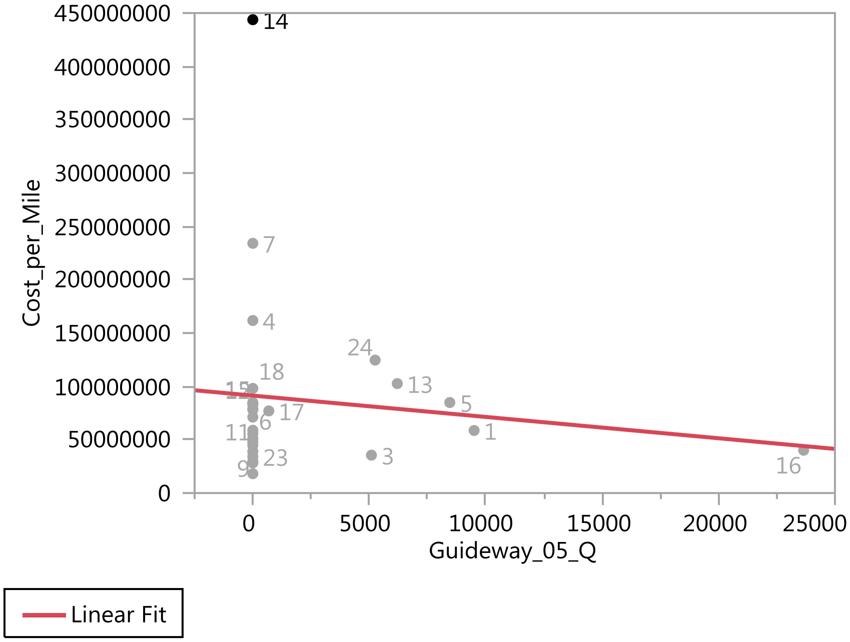
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	4.6266e+15	4.627e+15	0.6398
Error	25	1.8077e+17	7.231e+15	**Prob > F**
C. Total	26	1.854e+17		0.4313

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	76165948	21254021	3.58	0.0014*
Guideway_04_$	0.6914944	0.86448	0.80	0.4313

Bivariate Fit of Cost_per_Mile By Guideway_05_Q



Linear Fit

Cost_per_Mile = 91371650 - 1999.1517*Guideway_05_Q

Summary of Fit

RSquare	0.014779
RSquare Adj	-0.02463
Root Mean Square Error	85477412
Mean of Response	87013944
Observations (or Sum Wgts)	27

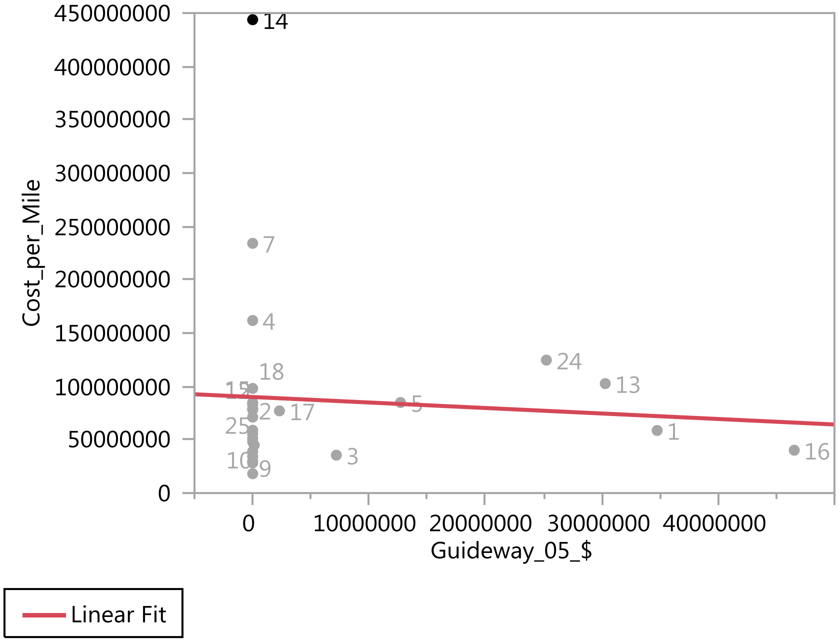
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	2.7401e+15	2.74e+15	0.3750
Error	25	1.8266e+17	7.306e+15	**Prob > F**
C. Total	26	1.854e+17		0.5458

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	91371650	17923228	5.10	<.0001*
Guideway_05_Q	-1999.152	3264.478	-0.61	0.5458

Bivariate Fit of Cost_per_Mile By Guideway_05_$



Linear Fit

Cost_per_Mile = 90052768 - 0.5154341*Guideway_05_$

Summary of Fit

RSquare	0.006033
RSquare Adj	-0.03373
Root Mean Square Error	85855996
Mean of Response	87013944
Observations (or Sum Wgts)	27

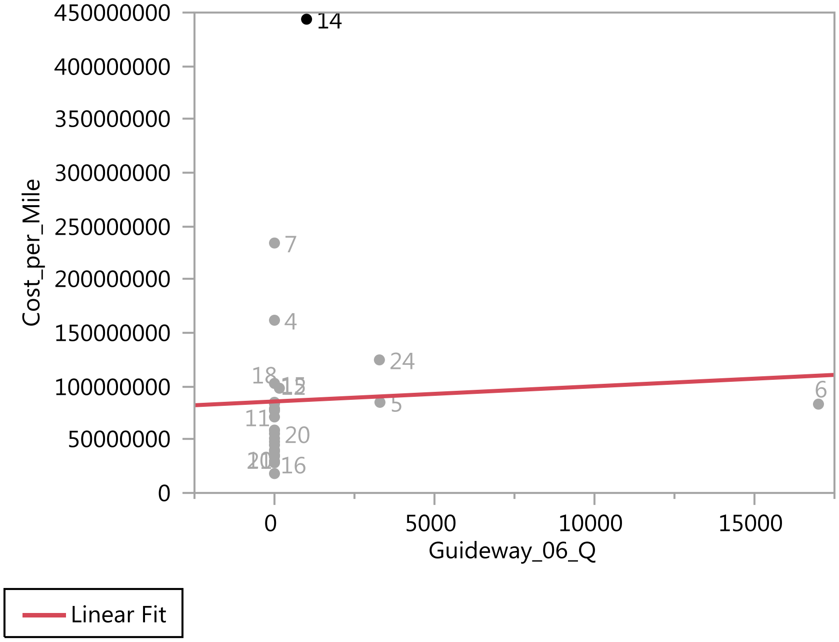
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1.1185e+15	1.118e+15	0.1517
Error	25	1.8428e+17	7.371e+15	**Prob > F**
C. Total	26	1.854e+17		0.7002

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	90052768	18272033	4.93	<.0001*
Guideway_05_$	-0.515434	1.323201	-0.39	0.7002

Bivariate Fit of Cost_per_Mile By Guideway_06_Q



Linear Fit

Cost_per_Mile = 85711838 + 1421.57*Guideway_06_Q

Summary of Fit

RSquare	0.00315
RSquare Adj	-0.03672
Root Mean Square Error	85980413
Mean of Response	87013944
Observations (or Sum Wgts)	27

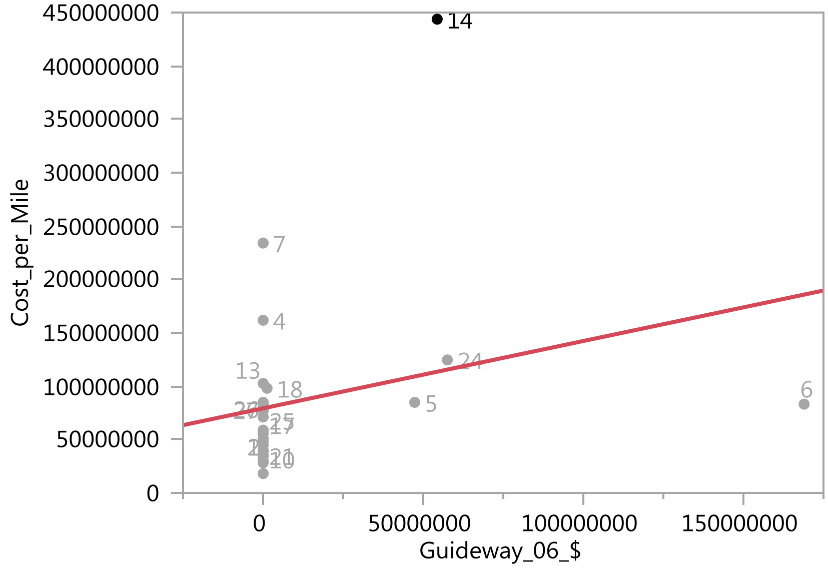
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	5.8401e+14	5.84e+14	0.0790
Error	25	1.8482e+17	7.393e+15	**Prob > F**
C. Total	26	1.854e+17		0.7810

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	85711838	17183221	4.99	<.0001*
Guideway_06_Q	1421.57	5057.74	0.28	0.7810

Bivariate Fit of Cost_per_Mile By Guideway_06_$



Linear Fit

Cost_per_Mile = 79322886 + 0.6303607*Guideway_06_$

Summary of Fit

RSquare	0.070806
RSquare Adj	0.033638
Root Mean Square Error	83011419
Mean of Response	87013944
Observations (or Sum Wgts)	27

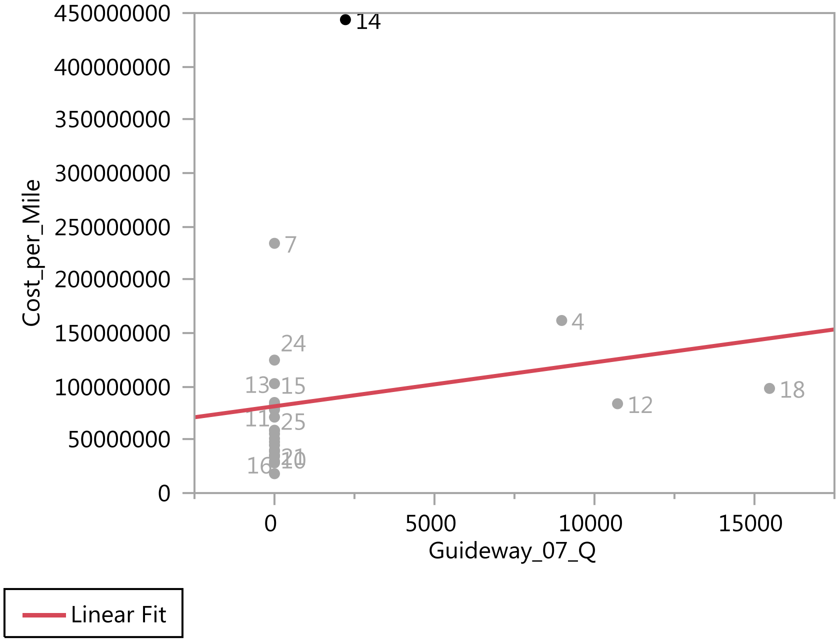
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1.3127e+16	1.313e+16	1.9050
Error	25	1.7227e+17	6.891e+15	**Prob > F**
C. Total	26	1.854e+17		0.1797

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	79322886	16919482	4.69	<.0001*
Guideway_06_$	0.6303607	0.456707	1.38	0.1797

Bivariate Fit of Cost_per_Mile By Guideway_07_Q



Linear Fit

Cost_per_Mile = 81340816 + 4117.894*Guideway_07_Q

Summary of Fit

RSquare	0.035128
RSquare Adj	-0.00682
Root Mean Square Error	88165069
Mean of Response	87498715
Observations (or Sum Wgts)	25

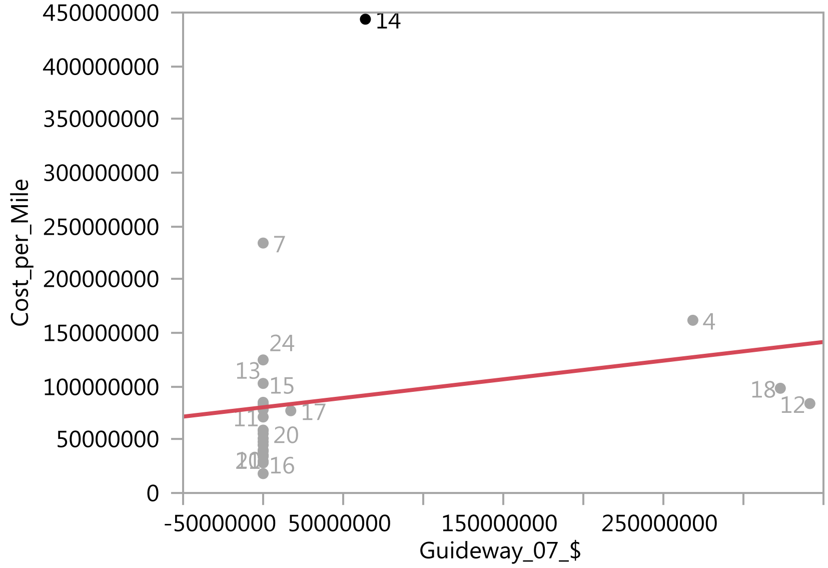
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	6.5089e+15	6.509e+15	0.8374
Error	23	1.7878e+17	7.773e+15	**Prob > F**
C. Total	24	1.8529e+17		0.3696

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	81340816	18873477	4.31	0.0003*
Guideway_07_Q	4117.894	4500.073	0.92	0.3696

Bivariate Fit of Cost_per_Mile By Guideway_07_$



Linear Fit

Cost_per_Mile = 80280037 + 0.1747522*Guideway_07_$

Summary of Fit

RSquare	0.042474
RSquare Adj	0.002577
Root Mean Square Error	86004034
Mean of Response	87096046
Observations (or Sum Wgts)	26

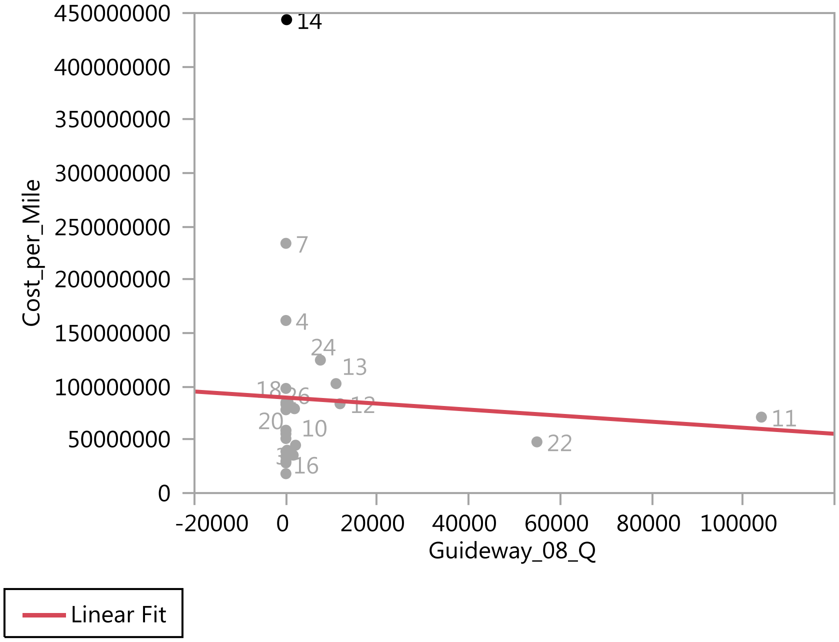
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	7.8744e+15	7.874e+15	1.0646
Error	24	1.7752e+17	7.397e+15	**Prob > F**
C. Total	25	1.854e+17		0.3125

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	80280037	18114295	4.43	0.0002*
Guideway_07_$	0.1747522	0.169368	1.03	0.3125

Bivariate Fit of Cost_per_Mile By Guideway_08_Q



Linear Fit

Cost_per_Mile = 89528926 - 283.03334*Guideway_08_Q

Summary of Fit

RSquare	0.005485
RSquare Adj	-0.03595
Root Mean Square Error	87626063
Mean of Response	87397968
Observations (or Sum Wgts)	26

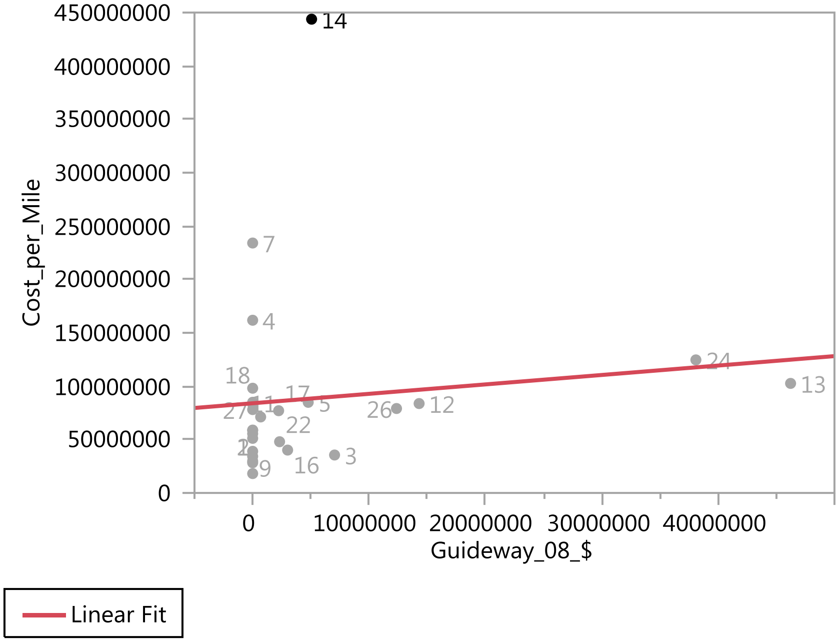
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1.0164e+15	1.016e+15	0.1324
Error	24	1.8428e+17	7.678e+15	**Prob > F**
C. Total	25	1.853e+17		0.7192

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	89528926	18155546	4.93	<.0001*
Guideway_08_Q	-283.0333	777.9154	-0.36	0.7192

Bivariate Fit of Cost_per_Mile By Guideway_08_$



Linear Fit

Cost_per_Mile = 84011775 + 0.8831444*Guideway_08_$

Summary of Fit

RSquare	0.014238
RSquare Adj	-0.02684
Root Mean Square Error	86828879
Mean of Response	88634807
Observations (or Sum Wgts)	26

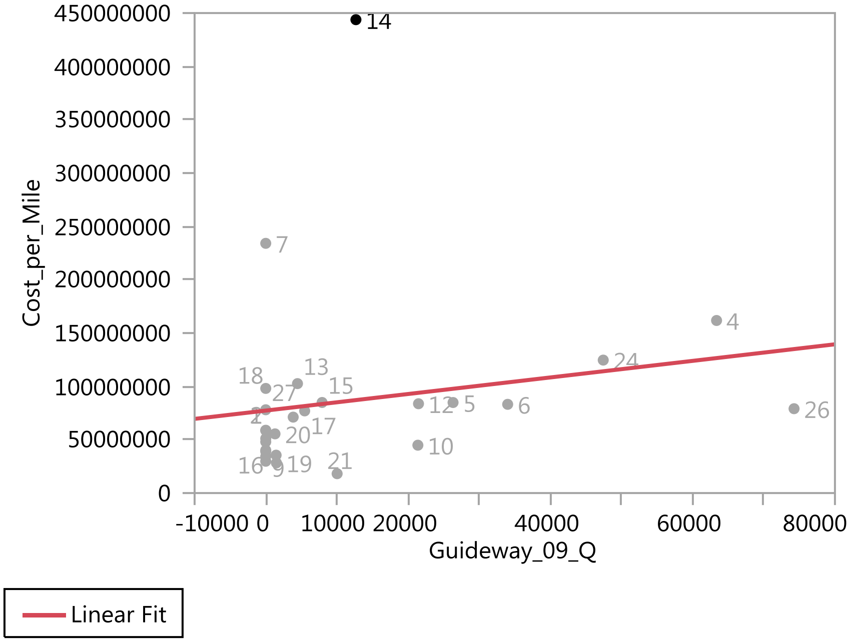
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	2.6134e+15	2.613e+15	0.3466
Error	24	1.8094e+17	7.539e+15	**Prob > F**
C. Total	25	1.8356e+17		0.5615

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	84011775	18751730	4.48	0.0002*
Guideway_08_$	0.8831444	1.500005	0.59	0.5615

Bivariate Fit of Cost_per_Mile By Guideway_09_Q



Linear Fit

Cost_per_Mile = 77321743 + 776.79386*Guideway_09_Q

Summary of Fit

RSquare	0.034941
RSquare Adj	-0.00366
Root Mean Square Error	84598303
Mean of Response	87013944
Observations (or Sum Wgts)	27

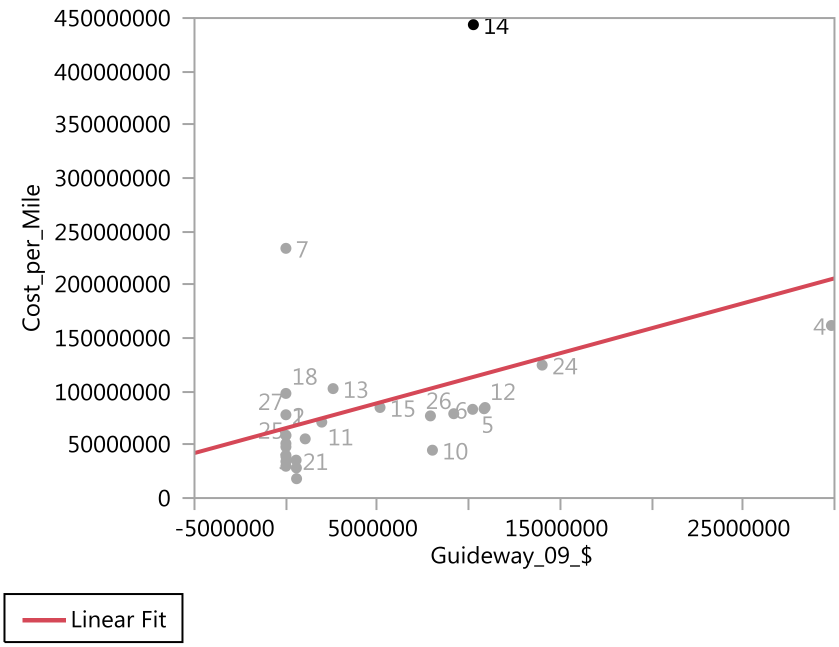
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	6.478e+15	6.478e+15	0.9051
Error	25	1.7892e+17	7.157e+15	**Prob > F**
C. Total	26	1.854e+17		0.3505

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	77321743	19205545	4.03	0.0005*
Guideway_09_Q	776.79386	816.4841	0.95	0.3505

Bivariate Fit of Cost_per_Mile By Guideway_09_$



Linear Fit

Cost_per_Mile = 65592412 + 4.6795753*Guideway_09_$

Summary of Fit

RSquare	0.143649
RSquare Adj	0.109395
Root Mean Square Error	79691232
Mean of Response	87013944
Observations (or Sum Wgts)	27

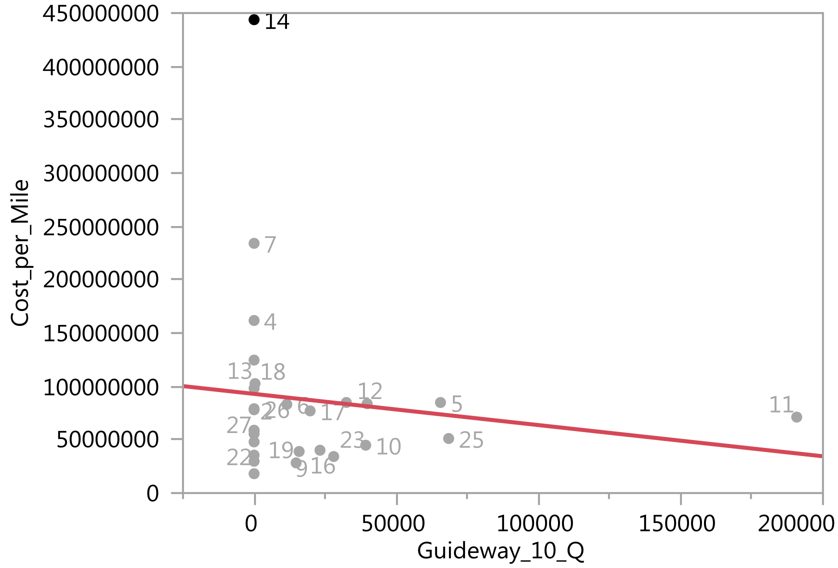
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	2.6632e+16	2.663e+16	4.1936
Error	25	1.5877e+17	6.351e+15	**Prob > F**
C. Total	26	1.854e+17		0.0512

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	65592412	18564328	3.53	0.0016*
Guideway_09_$	4.6795753	2.285131	2.05	0.0512

Bivariate Fit of Cost_per_Mile By Guideway_10_Q



Linear Fit

Linear Fit

Cost_per_Mile = 92964290 - 292.41151*Guideway_10_Q

Summary of Fit

RSquare	0.018755
RSquare Adj	-0.02049
Root Mean Square Error	85304774
Mean of Response	87013944
Observations (or Sum Wgts)	27

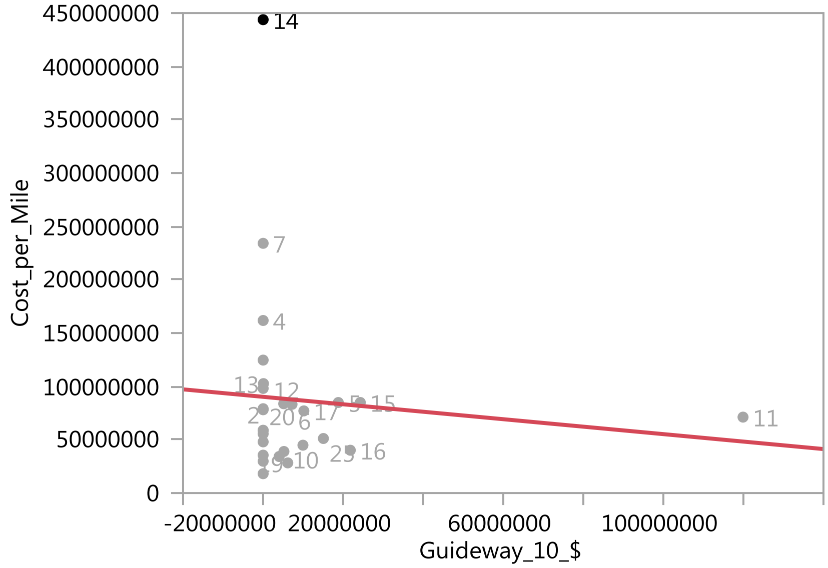
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	3.4772e+15	3.477e+15	0.4778
Error	25	1.8192e+17	7.277e+15	**Prob > F**
C. Total	26	1.854e+17		0.4958

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	92964290	18536785	5.02	<.0001*
Guideway_10_Q	-292.4115	423.013	-0.69	0.4958

Bivariate Fit of Cost_per_Mile By Guideway_10_$



Linear Fit

Linear Fit

Cost_per_Mile = 90222387 - 0.3497681*Guideway_10_$

Summary of Fit

RSquare	0.00931
RSquare Adj	-0.03032
Root Mean Square Error	85714362
Mean of Response	87013944
Observations (or Sum Wgts)	27

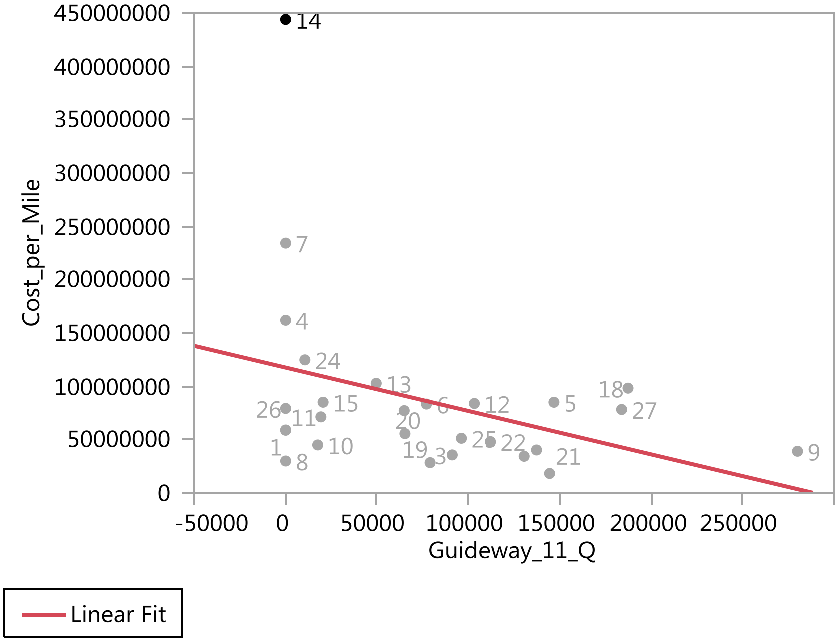
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1.726e+15	1.726e+15	0.2349
Error	25	1.8367e+17	7.347e+15	**Prob > F**
C. Total	26	1.854e+17		0.6321

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	90222387	17774349	5.08	<.0001*
Guideway_10_$	-0.349768	0.721627	-0.48	0.6321

Bivariate Fit of Cost_per_Mile By Guideway_11_Q



Linear Fit

Cost_per_Mile = 117398776 - 407.11601*Guideway_11_Q

Summary of Fit

RSquare	0.125108
RSquare Adj	0.090112
Root Mean Square Error	80549336
Mean of Response	87013944
Observations (or Sum Wgts)	27

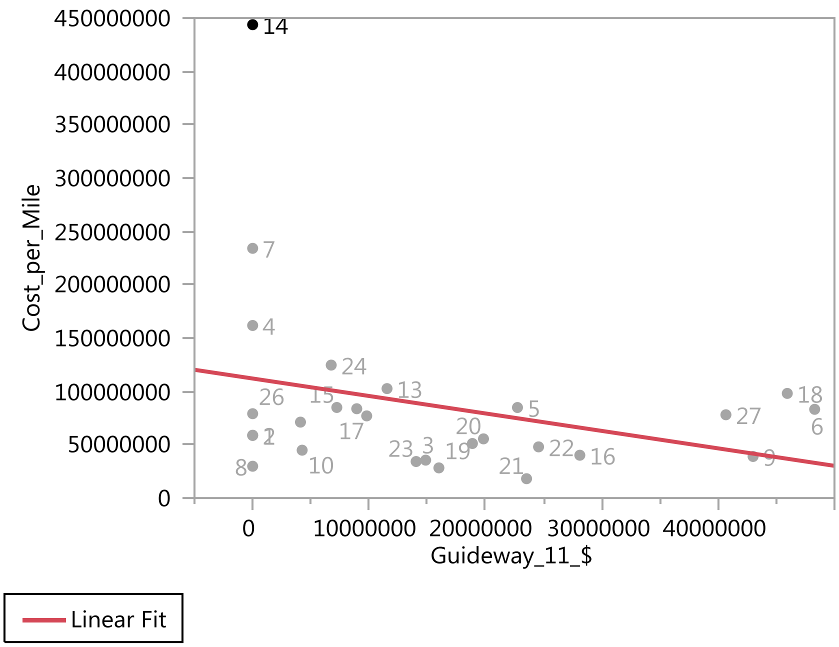
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	2.3195e+16	2.319e+16	3.5749
Error	25	1.622e+17	6.488e+15	**Prob > F**
C. Total	26	1.854e+17		0.0703

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	117398776	22328373	5.26	<.0001*
Guideway_11_Q	-407.116	215.3197	-1.89	0.0703

Bivariate Fit of Cost_per_Mile By Guideway_11_$



Linear Fit

Cost_per_Mile = 112064965 - 1.6370126*Guideway_11_$

Summary of Fit

RSquare	0.086109
RSquare Adj	0.049553
Root Mean Square Error	82325034
Mean of Response	87013944
Observations (or Sum Wgts)	27

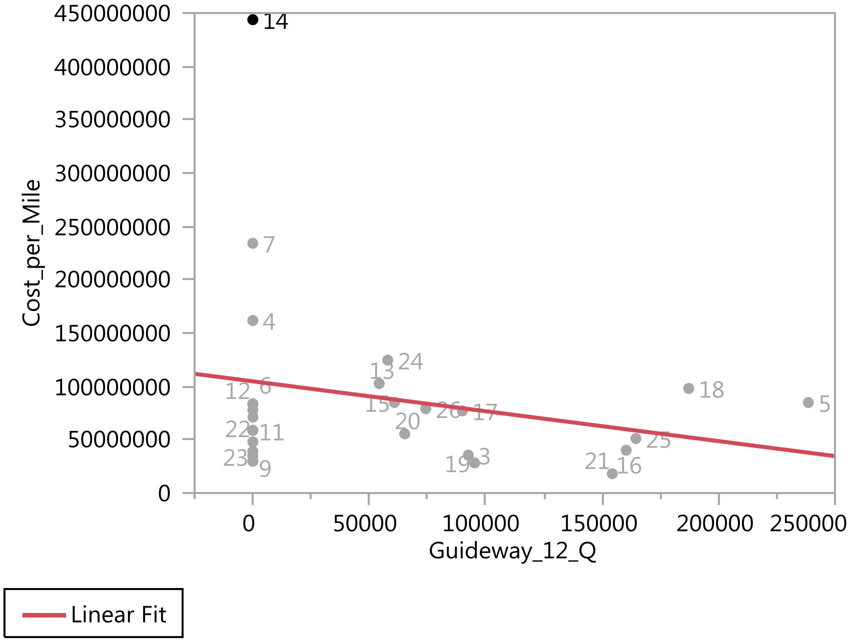
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1.5965e+16	1.596e+16	2.3555
Error	25	1.6944e+17	6.777e+15	**Prob > F**
C. Total	26	1.854e+17		0.1374

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	112064965	22747095	4.93	<.0001*
Guideway_11_$	-1.637013	1.066611	-1.53	0.1374

Bivariate Fit of Cost_per_Mile By Guideway_12_Q



Linear Fit

Cost_per_Mile = 104762139 - 280.3915*Guideway_12_Q

Summary of Fit

RSquare	0.054853
RSquare Adj	0.015472
Root Mean Square Error	85021292
Mean of Response	88634807
Observations (or Sum Wgts)	26

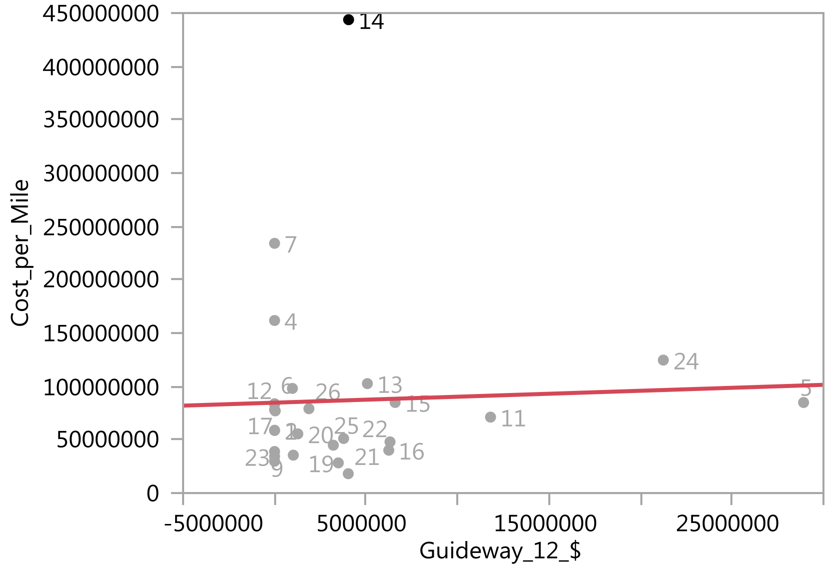
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1.0069e+16	1.007e+16	1.3929
Error	24	1.7349e+17	7.229e+15	**Prob > F**
C. Total	25	1.8356e+17		0.2495

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	104762139	21558113	4.86	<.0001*
Guideway_12_Q	-280.3915	237.5787	-1.18	0.2495

Bivariate Fit of Cost_per_Mile By Guideway_12_$



Linear Fit

Cost_per_Mile = 84744197 + 0.5576408*Guideway_12_$

Summary of Fit

RSquare	0.002019
RSquare Adj	-0.0379
Root Mean Square Error	86029178
Mean of Response	87013944
Observations (or Sum Wgts)	27

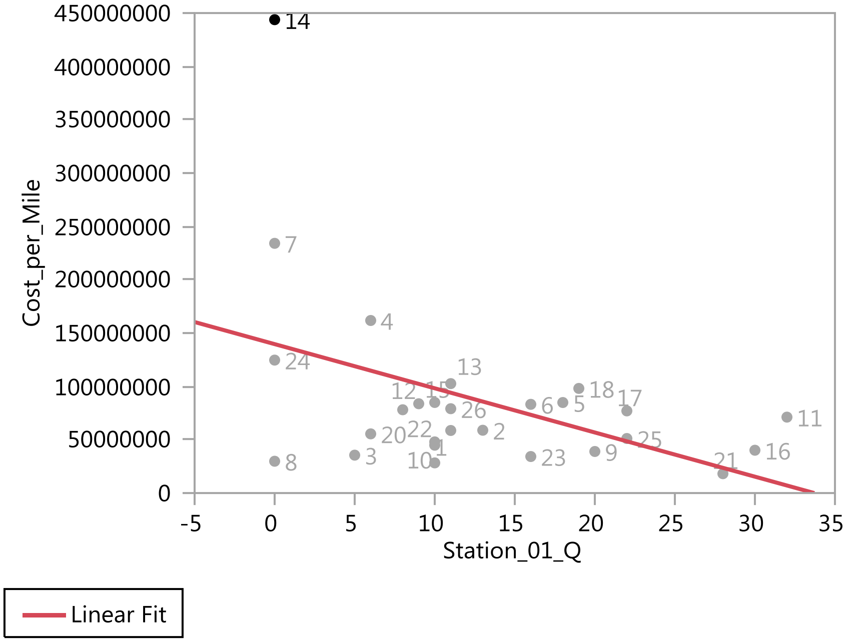
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	3.7432e+14	3.743e+14	0.0506
Error	25	1.8503e+17	7.401e+15	**Prob > F**
C. Total	26	1.854e+17		0.8239

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	84744197	19390023	4.37	0.0002*
Guideway_12_$	0.5576408	2.4796	0.22	0.8239

Bivariate Fit of Cost_per_Mile By Station_01_Q



Linear Fit

Cost_per_Mile = 139633633 - 4142074.7*Station_01_Q

Summary of Fit

RSquare	0.193373
RSquare Adj	0.161108
Root Mean Square Error	77342997
Mean of Response	87013944
Observations (or Sum Wgts)	27

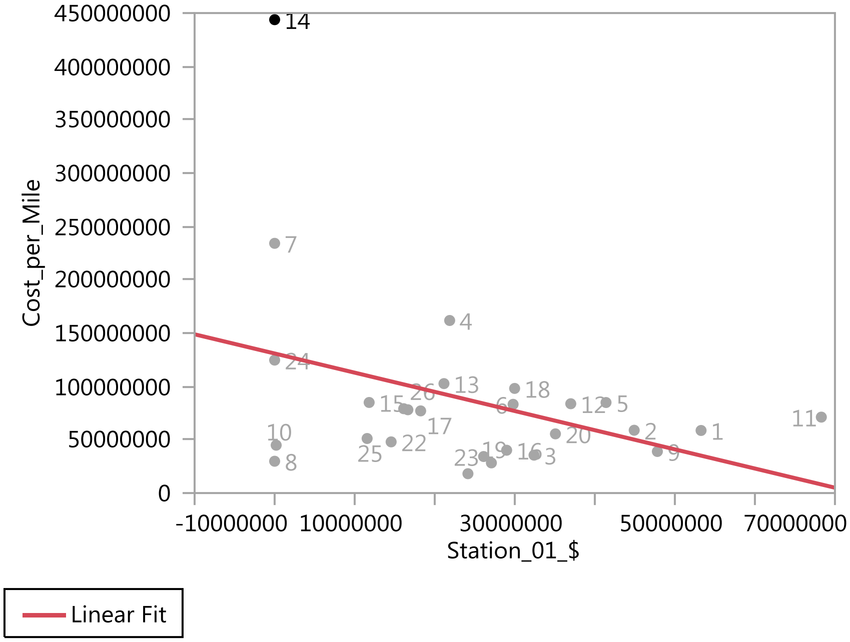
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	3.5851e+16	3.585e+16	5.9933
Error	25	1.4955e+17	5.982e+15	**Prob > F**
C. Total	26	1.854e+17		0.0217*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	139633633	26144678	5.34	<.0001*
Station_01_Q	-4142075	1691945	-2.45	0.0217*

Bivariate Fit of Cost_per_Mile By Station_01_$



Linear Fit

Cost_per_Mile = 130850492 - 1.7976776*Station_01_$

Summary of Fit

RSquare	0.139762
RSquare Adj	0.105352
Root Mean Square Error	79871888
Mean of Response	87013944
Observations (or Sum Wgts)	27

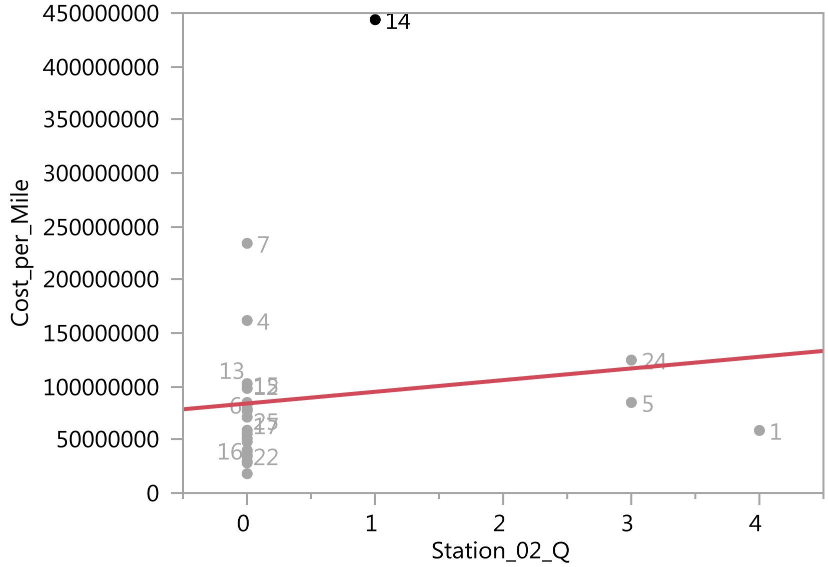
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	2.5912e+16	2.591e+16	4.0617
Error	25	1.5949e+17	6.38e+15	**Prob > F**
C. Total	26	1.854e+17		0.0547

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	130850492	26634352	4.91	<.0001*
Station_01_$	-1.797678	0.891983	-2.02	0.0547

Bivariate Fit of Cost_per_Mile By Station_02_Q



Linear Fit

Cost_per_Mile = 84005157 + 10942811*Station_02_Q

Summary of Fit

RSquare	0.019797
RSquare Adj	-0.02105
Root Mean Square Error	86583703
Mean of Response	88634807
Observations (or Sum Wgts)	26

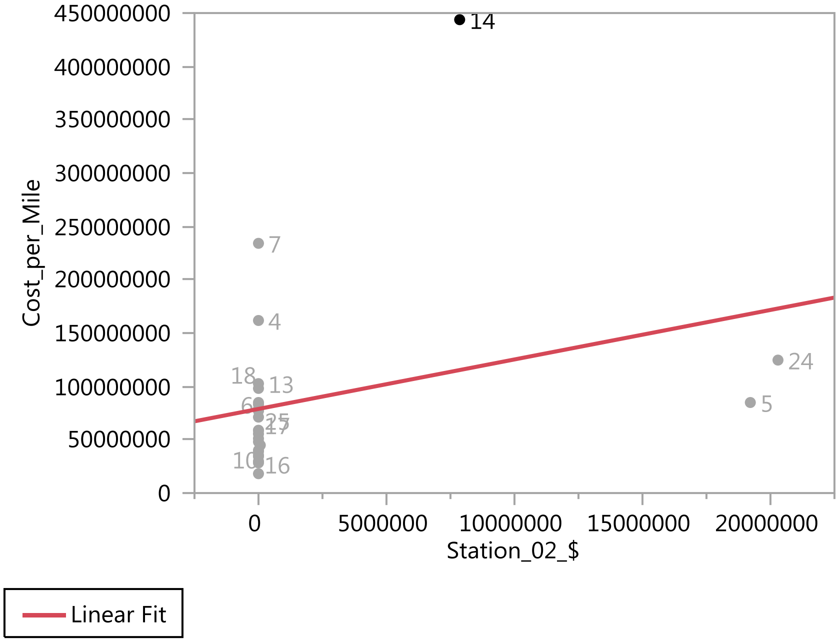
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	3.6338e+15	3.634e+15	0.4847
Error	24	1.7992e+17	7.497e+15	**Prob > F**
C. Total	25	1.8356e+17		0.4930

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	84005157	18236088	4.61	0.0001*
Station_02_Q	10942811	15717531	0.70	0.4930

Bivariate Fit of Cost_per_Mile By Station_02_$



Linear Fit

Cost_per_Mile = 78861972 + 4.6399611*Station_02_$

Summary of Fit

RSquare	0.088184
RSquare Adj	0.051711
Root Mean Square Error	82231521
Mean of Response	87013944
Observations (or Sum Wgts)	27

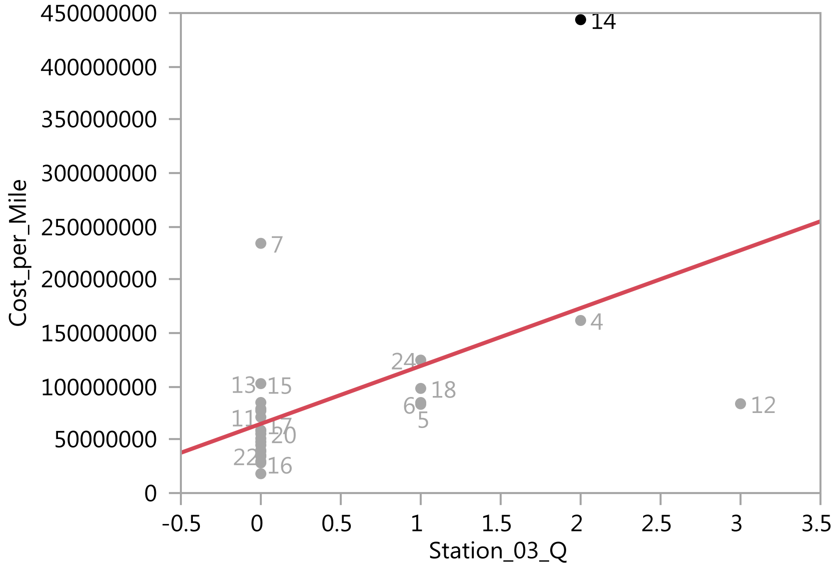
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1.6349e+16	1.635e+16	2.4178
Error	25	1.6905e+17	6.762e+15	**Prob > F**
C. Total	26	1.854e+17		0.1325

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	78861972	16671259	4.73	<.0001*
Station_02_$	4.6399611	2.984036	1.55	0.1325

Bivariate Fit of Cost_per_Mile By Station_03_Q



Linear Fit

Linear Fit

Cost_per_Mile = 64915068 + 54242695*Station_03_Q

Summary of Fit

RSquare	0.262147
RSquare Adj	0.232633
Root Mean Square Error	73972389
Mean of Response	87013944
Observations (or Sum Wgts)	27

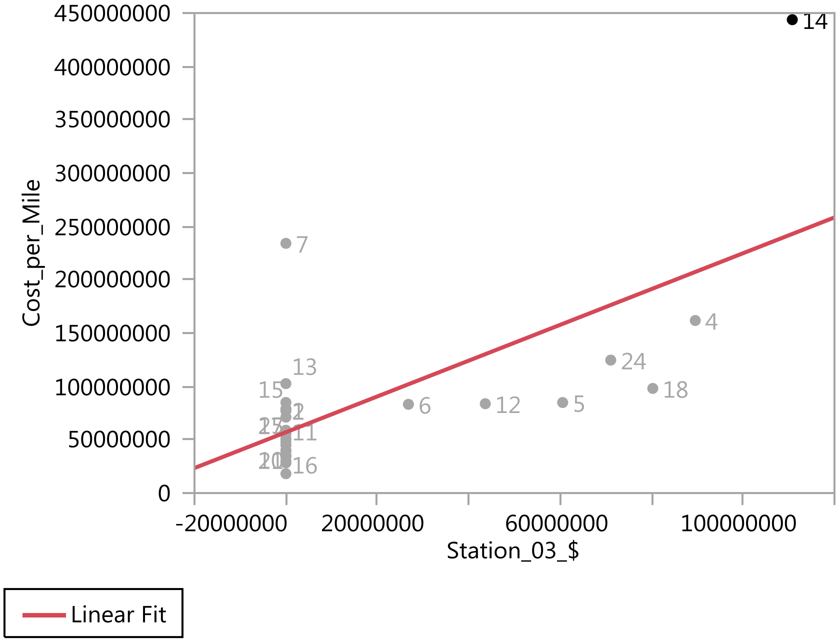
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	4.8602e+16	4.86e+16	8.8821
Error	25	1.368e+17	5.472e+15	**Prob > F**
C. Total	26	1.854e+17		0.0063*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	64915068	16051361	4.04	0.0004*
Station_03_Q	54242695	18200533	2.98	0.0063*

Bivariate Fit of Cost_per_Mile By Station_03_$



Linear Fit

Cost_per_Mile = 57018607 + 1.6782131*Station_03_$

Summary of Fit

RSquare	0.447207
RSquare Adj	0.425095
Root Mean Square Error	64027405
Mean of Response	87013944
Observations (or Sum Wgts)	27

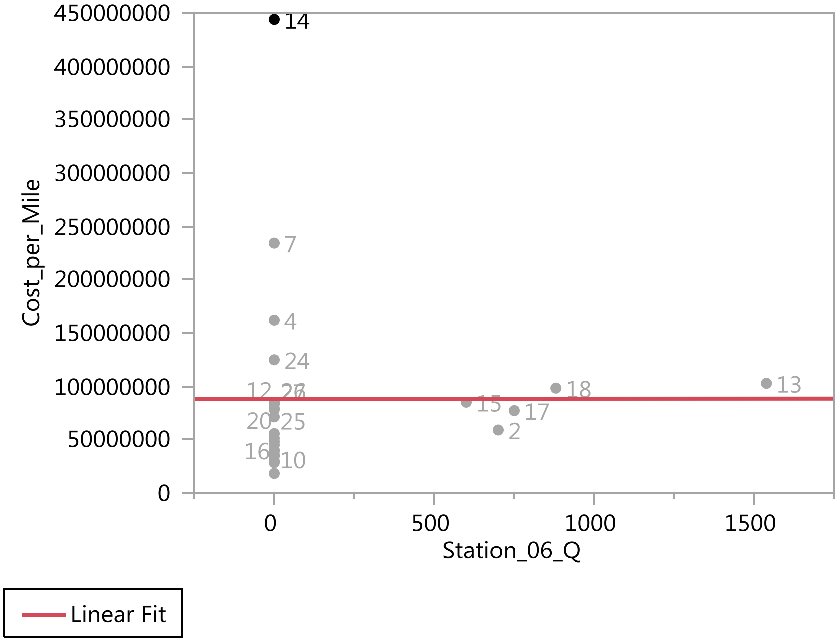
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	8.2912e+16	8.291e+16	20.2249
Error	25	1.0249e+17	4.1e+15	**Prob > F**
C. Total	26	1.854e+17		0.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	57018607	14011404	4.07	0.0004*
Station_03_$	1.6782131	0.373168	4.50	0.0001*

Bivariate Fit of Cost_per_Mile By Station_06_Q



Linear Fit

Cost_per_Mile = 88084205 + 114.75627*Station_06_Q

Summary of Fit

RSquare	2.7e-7
RSquare Adj	-0.04167
Root Mean Square Error	87694007
Mean of Response	88103925
Observations (or Sum Wgts)	26

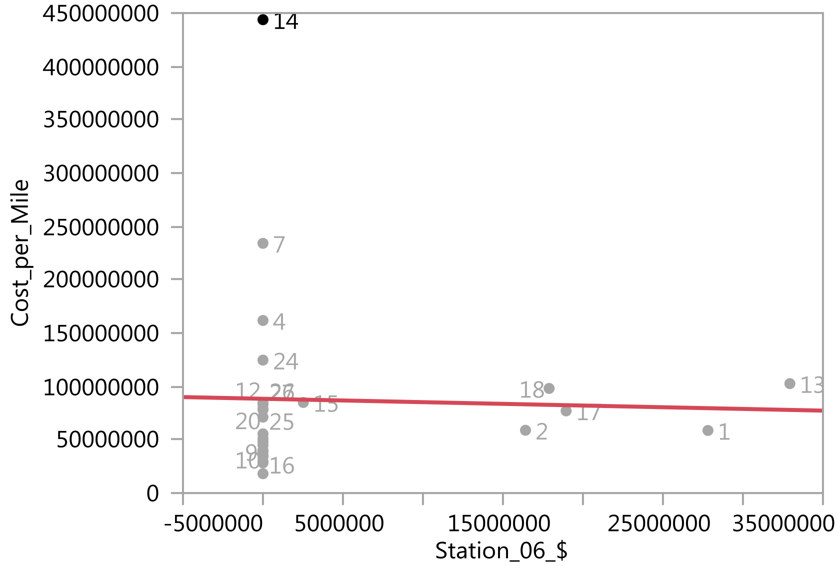
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	4.9839e+10	4.984e+10	0.0000
Error	24	1.8457e+17	7.69e+15	**Prob > F**
C. Total	25	1.8457e+17		0.9980

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	88084205	18862293	4.67	<.0001*
Station_06_Q	114.75627	45077.9	0.00	0.9980

Bivariate Fit of Cost_per_Mile By Station_06_$



Linear Fit

Cost_per_Mile = 88381873 - 0.3170363*Station_06_$

Summary of Fit

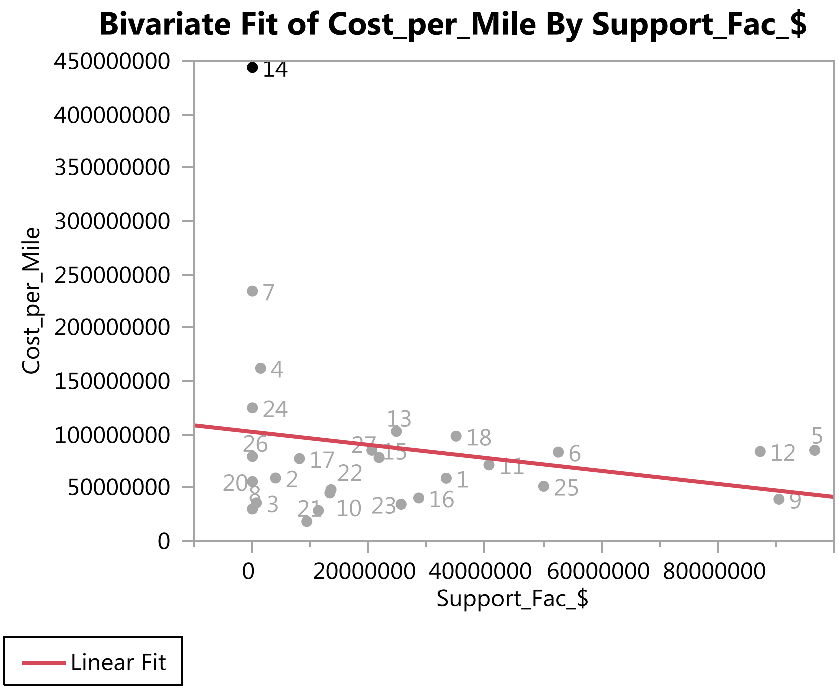
RSquare	0.001252
RSquare Adj	-0.0387
Root Mean Square Error	86062225
Mean of Response	87013944
Observations (or Sum Wgts)	27

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	2.3214e+14	2.321e+14	0.0313
Error	25	1.8517e+17	7.407e+15	**Prob > F**
C. Total	26	1.854e+17		0.8609

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	88381873	18276408	4.84	<.0001*
Station_06_$	-0.317036	1.790809	-0.18	0.8609



Linear Fit

Cost_per_Mile = 102124945 - 0.609935*Support_Fac_$

Summary of Fit

RSquare	0.043087
RSquare Adj	0.00481
Root Mean Square Error	84240484
Mean of Response	87013944
Observations (or Sum Wgts)	27

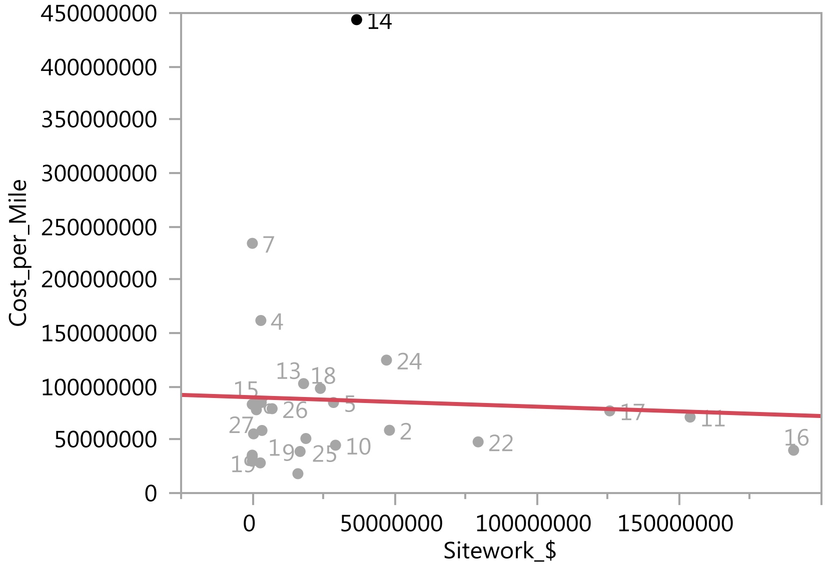
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	7.9883e+15	7.988e+15	1.1257
Error	25	1.7741e+17	7.096e+15	**Prob > F**
C. Total	26	1.854e+17		0.2988

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	102124945	21579637	4.73	<.0001*
Support_Fac_$	-0.609935	0.574879	-1.06	0.2988

Bivariate Fit of Cost_per_Mile By Sitework_$



Linear Fit

Linear Fit

Cost_per_Mile = 89803555 - 0.0880209*Sitework_$

Summary of Fit

RSquare	0.002685
RSquare Adj	-0.03721
Root Mean Square Error	86000447
Mean of Response	87013944
Observations (or Sum Wgts)	27

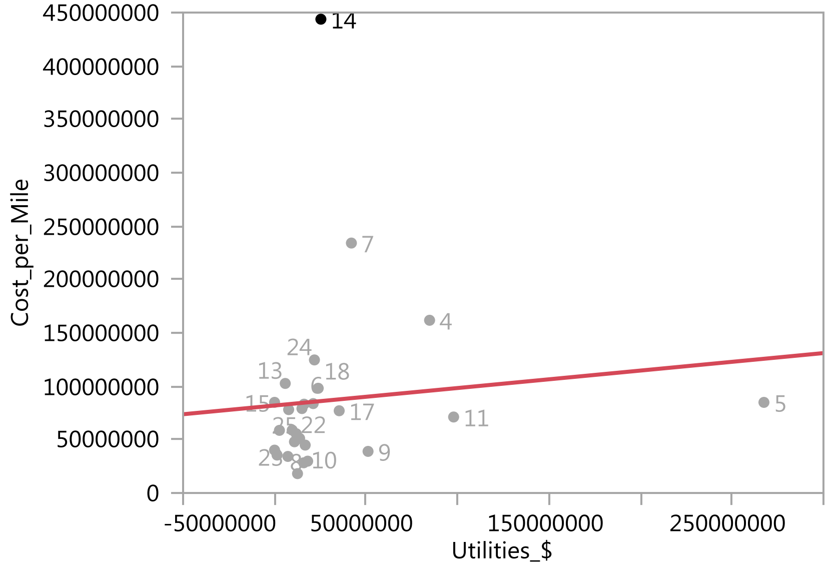
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	4.9788e+14	4.979e+14	0.0673
Error	25	1.849e+17	7.396e+15	**Prob > F**
C. Total	26	1.854e+17		0.7974

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	89803555	19736540	4.55	0.0001*
Sitework_$	-0.088021	0.339254	-0.26	0.7974

Bivariate Fit of Cost_per_Mile By Utilities_$



Linear Fit

Linear Fit

Cost_per_Mile = 81948673 + 0.1636484*Utilities_$

Summary of Fit

RSquare	0.010453
RSquare Adj	-0.02913
Root Mean Square Error	85664871
Mean of Response	87013944
Observations (or Sum Wgts)	27

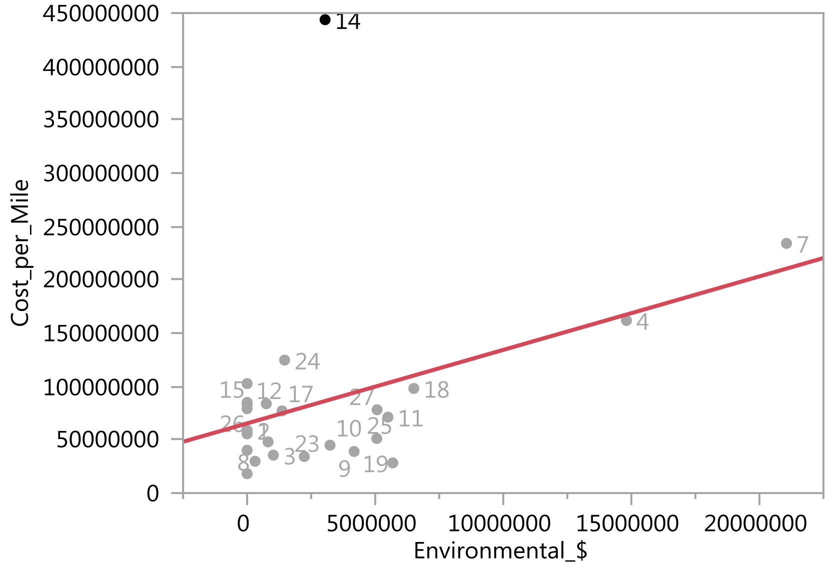
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1.938e+15	1.938e+15	0.2641
Error	25	1.8346e+17	7.338e+15	**Prob > F**
C. Total	26	1.854e+17		0.6118

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	81948673	19207969	4.27	0.0002*
Utilities_$	0.1636484	0.318444	0.51	0.6118

Bivariate Fit of Cost_per_Mile By Environmental_$



Linear Fit

Linear Fit

Cost_per_Mile = 65340496 + 6.8896184*Environmental_$

Summary of Fit

RSquare	0.155713
RSquare Adj	0.120534
Root Mean Square Error	80758570
Mean of Response	87096046
Observations (or Sum Wgts)	26

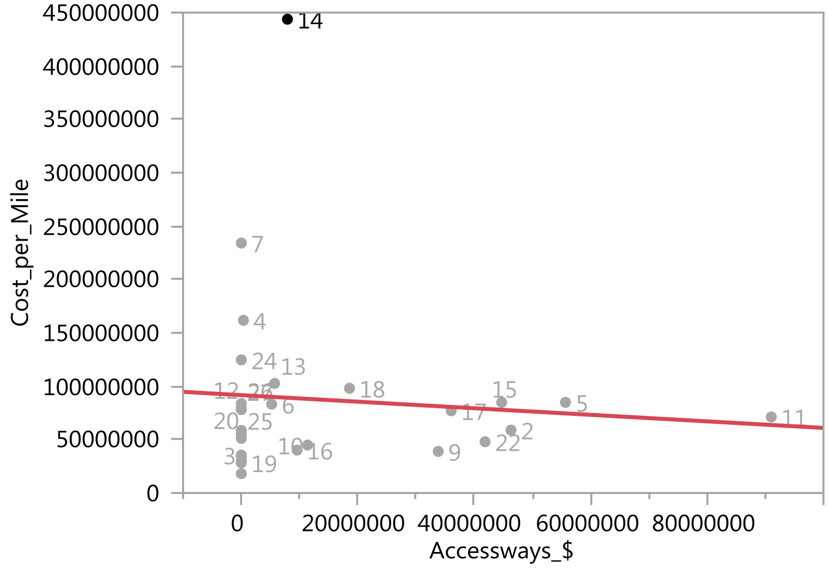
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	2.8868e+16	2.887e+16	4.4263
Error	24	1.5653e+17	6.522e+15	**Prob > F**
C. Total	25	1.854e+17		0.0461*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	65340496	18914891	3.45	0.0021*
Environmental_$	6.8896184	3.27471	2.10	0.0461*

Bivariate Fit of Cost_per_Mile By Accessways_$



Linear Fit

Linear Fit

Cost_per_Mile = 91656830 - 0.3070338*Accessways_$

Summary of Fit

RSquare	0.007245
RSquare Adj	-0.03247
Root Mean Square Error	85803651
Mean of Response	87013944
Observations (or Sum Wgts)	27

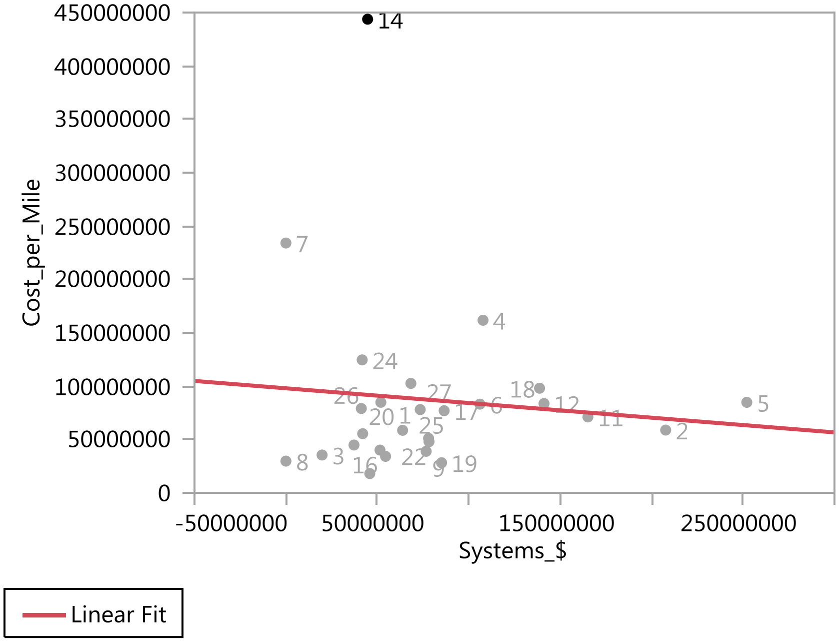
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1.3431e+15	1.343e+15	0.1824
Error	25	1.8406e+17	7.362e+15	**Prob > F**
C. Total	26	1.854e+17		0.6729

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	91656830	19769563	4.64	<.0001*
Accessways_$	-0.307034	0.718839	-0.43	0.6729

Bivariate Fit of Cost_per_Mile By Systems_$



Linear Fit

Cost_per_Mile = 98025820 - 0.1376904*Systems_$

Summary of Fit

RSquare	0.009169
RSquare Adj	-0.03046
Root Mean Square Error	85720466
Mean of Response	87013944
Observations (or Sum Wgts)	27

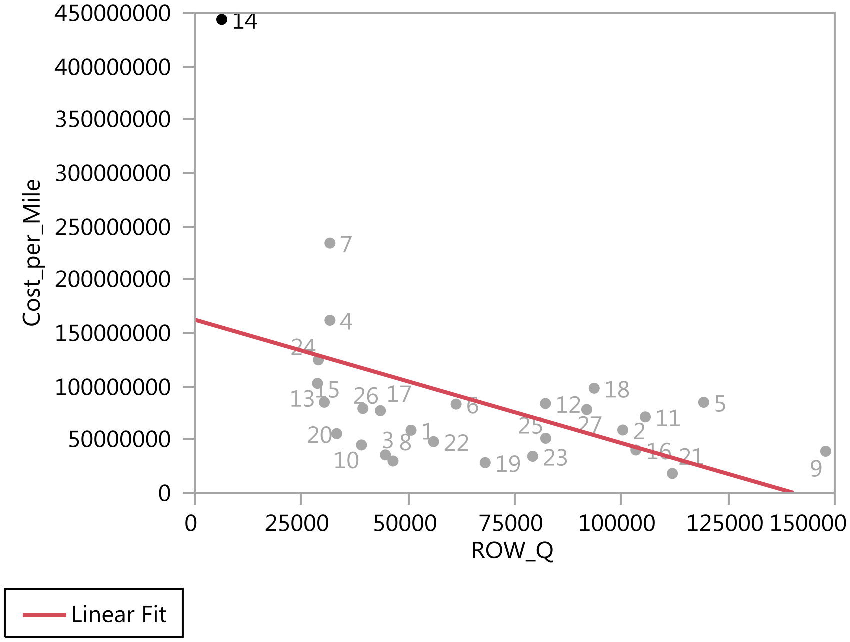
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1.6998e+15	1.7e+15	0.2313
Error	25	1.837e+17	7.348e+15	**Prob > F**
C. Total	26	1.854e+17		0.6347

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	98025820	28219337	3.47	0.0019*
Systems_$	-0.13769	0.286275	-0.48	0.6347

Bivariate Fit of Cost_per_Mile By ROW_Q



Linear Fit

Cost_per_Mile = 162297701 - 1156.4378*ROW_Q

Summary of Fit

RSquare	0.231254
RSquare Adj	0.200504
Root Mean Square Error	75505059
Mean of Response	87013944
Observations (or Sum Wgts)	27

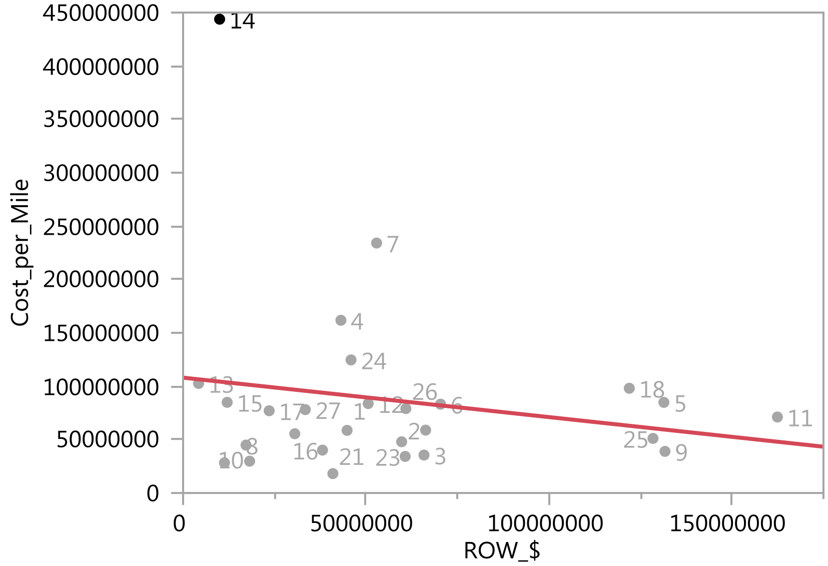
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	4.2874e+16	4.287e+16	7.5205
Error	25	1.4253e+17	5.701e+15	**Prob > F**
C. Total	26	1.854e+17		0.0111*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	162297701	31060828	5.23	<.0001*
ROW_Q	-1156.438	421.6956	-2.74	0.0111*

Bivariate Fit of Cost_per_Mile By ROW_$



Linear Fit

Cost_per_Mile = 108041964 - 0.3695581*ROW_$

Summary of Fit

RSquare	0.035086
RSquare Adj	-0.00351
Root Mean Square Error	84591931
Mean of Response	87013944
Observations (or Sum Wgts)	27

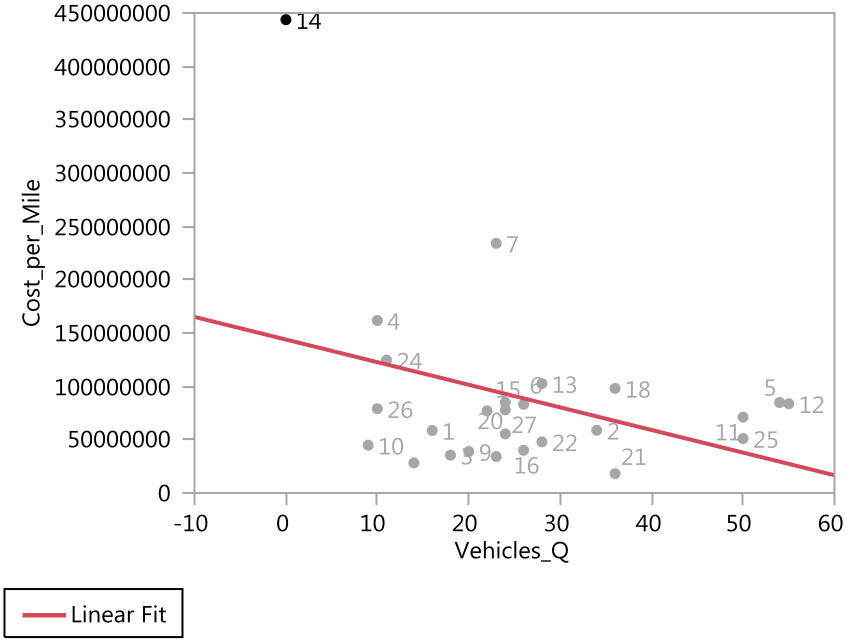
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	6.5049e+15	6.505e+15	0.9090
Error	25	1.7889e+17	7.156e+15	**Prob > F**
C. Total	26	1.854e+17		0.3495

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	108041964	27412591	3.94	0.0006*
ROW_$	-0.369558	0.387606	-0.95	0.3495

Bivariate Fit of Cost_per_Mile By Vehicles_Q



Linear Fit

Cost_per_Mile = 143929752 - 2120154.6*Vehicles_Q

Summary of Fit

RSquare	0.126947
RSquare Adj	0.090569
Root Mean Square Error	81368215
Mean of Response	89213454
Observations (or Sum Wgts)	26

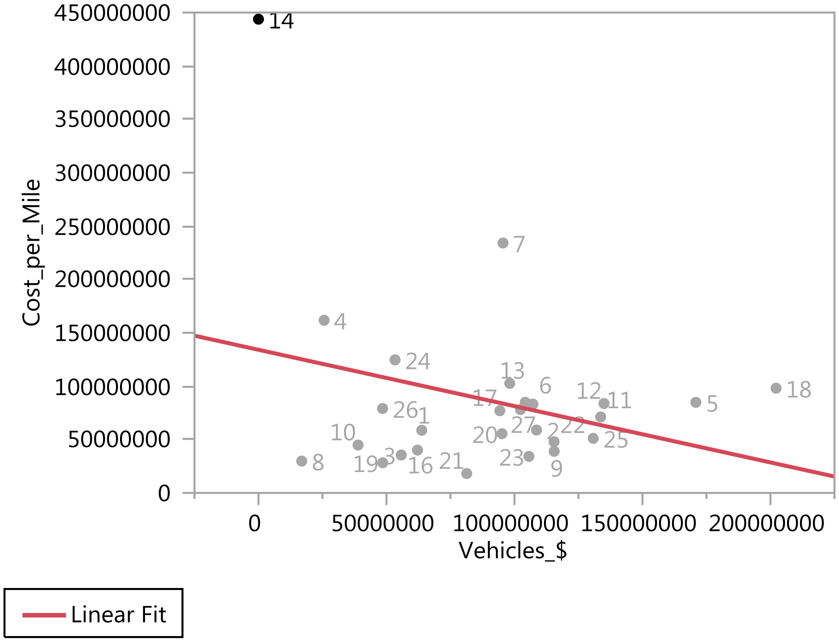
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	2.3105e+16	2.31e+16	3.4897
Error	24	1.589e+17	6.621e+15	**Prob > F**
C. Total	25	1.82e+17		0.0740

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	143929752	33354989	4.32	0.0002*
Vehicles_Q	-2120155	1134937	-1.87	0.0740

Bivariate Fit of Cost_per_Mile By Vehicles_$



Linear Fit

Cost_per_Mile = 134105556 - 0.5277422*Vehicles_$

Summary of Fit

RSquare	0.082733
RSquare Adj	0.046042
Root Mean Square Error	82476954
Mean of Response	87013944
Observations (or Sum Wgts)	27

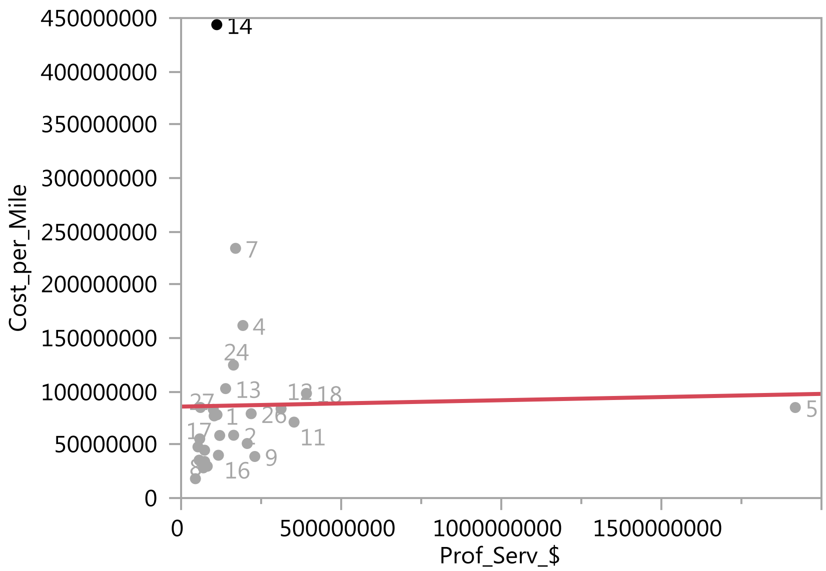
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1.5339e+16	1.534e+16	2.2549
Error	25	1.7006e+17	6.802e+15	**Prob > F**
C. Total	26	1.854e+17		0.1457

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	134105556	35148613	3.82	0.0008*
Vehicles_$	-0.527742	0.351448	-1.50	0.1457

Bivariate Fit of Cost_per_Mile By Prof_Serv_$



Linear Fit

Cost_per_Mile = 85757440 + 0.0059566*Prof_Serv_$

Summary of Fit

RSquare	0.000622
RSquare Adj	-0.03935
Root Mean Square Error	86089389
Mean of Response	87013944
Observations (or Sum Wgts)	27

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1.1523e+14	1.152e+14	0.0155
Error	25	1.8528e+17	7.411e+15	**Prob > F**
C. Total	26	1.854e+17		0.9018

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	85757440	19391796	4.42	0.0002*
Prof_Serv_$	0.0059566	0.047772	0.12	0.9018

BIBLIOGRAPHY

AECOM. (2010). TRCP Report 138: Estimating soft costs for major public transportation fixed guideway projects. Washington, DC: Transit Cooperative Research Program.

Akintola, A. (2000). Analysis of factors influencing project cost estimating practice. Construction Management and Economics, 18, 77 - 89.

American Public Transportation Association (APTA). (2005). *PUBLIC TRANSPORTATION FACT BOOK.* Washington, DC: American Public Transportation Association.

Assetts, D. L. (2013). White Paper: Mass Transit Capital Planning. Decision lens Proprietary Information.

Ayyub, B., & McCuen, R. (2003). Probability, statistics, and reliability for engineers and scientists. Boca Raton: Chapman & Hall/CRC Press LLC.

Booz - Allen & Hamilton (1991). Light Rail Transit Capital Cost Study. Washington, DC: US DOT, Urban Mass Transportation Administration.

Button, K. (2010). The accuracy of transit system ridership forecasts and capital cost estimates. International Journal of Transport Economics, Vol. XXXVII, No.2, 155 - 168.

Cantarelli,C.C., Flyvbjerg, B., Molin, E.J.E., and van Wee, B. (2010). Cost overruns in Large-Scale Transportation Infrastructure Projects: Explanations and Their Theoretical Embeddedness. *European Journal of Transport and Infrastructure Research*, 10 (1): 5-18.

Carr, R. (1989). Cost-estimating principles. *Journal of Construction Engineering and Management*, Vol. 115, No. 4, 545 - 551.

Chen, Y.-J., Li, Z.-C., & Lam, W. H. K. (2014). Modeling transit technology selection in a linear transportation corridor. *Journal of Advanced Transportation*, *49*(1), 48–72.

Collins, J. (2005). Management challenges unique to transit projects. 2005 AACE International Transactions, PM.15PM, PM.15.1 – PM.15.6

Cook, D. R., & Weisberg, S. (1994). *An introduction to regression graphics*. New York, NY: John Wiley & Sons.

Dantata, N., Touran, A. and Schneck, D. (2006). Trends in U.S. rail transit project cost overrun. TRB 2006 Annual Meeting.

ECONORTHWEST (2002). TCPR report 78: Estimating the benefits and costs of public transit projects: A Guidebook for practitioners. TRANSIT COOPERATIVE RESEARCH PROGRAM.

FHWA. (2016). Public Involvement Process. Retrieved from FHWA: http://www.fhwa.dot.gov/planning/public_involvement/

FHWA/FTA. (2015). A guide to transportation decision making. Washington, DC: USDOT/FHWA/FTA.

FHWA/FTA. (2016). Environmental impacts and related procedures. Retrieved from http://environment.transportation.org/environmental_topics/nepa_process

FHWA/FTA. (2016). Public involvement process. Retrieved from http://www.fhwa.dot.gov/planning/public_involvement

FHWA/FTA. (2016). Transportation Planning Capacity Building – Transit at the Table. Retrieved from https://www.planning.dot.gov/focus_transit_attable.asp

Flury, B., & Riedwyl, H. (1998). Multivariate statistics: a practical approach. London: Chapman and Hall Ltd.

Freemark, Y. (2016, January 6). Openings and Construction Start Planned for 2016. Retrieved from The Transport Politic: http://www.thetransportpolitic.com/2016/01/06/openings-and-construction-starts-planned-for-2016/

Freund, R., Littell, R., & Creighton, L. (2003). Regression using JMP. Cary: SAS Institute and Wiley.

Federal Transit Administration. (2016). Annual Report on Funding Recommendations, FY 2017, Capital Investment Grant Program. FTA.

Federal Transit Administration. (2016). Before-and-After Study. Retrieved from https://www.transit.dot.gov/funding/grant-programs/capital-investments/guidance-and-after-studies-new-starts-projects

Federal Transit Administration. (2016). Capital Cost Database. Retrieved from Federal Transit Administration: https://www.transit.dot.gov/funding/grant-programs/capital-investments/capital-cost-database

Federal Transit Administration. (2016). Capital Investment Grant (CIP) Program. Retrieved from Federal Transit Administration: https://www.transit.dot.gov/funding/grant-programs/capital-investments/capital-investment-grant-program

Federal Transit Administration (2003). Predicted and actual impacts of New Starts Projects — Capital cost, operating cost and ridership data. Washington, DC: USDOT/FTA.

Federal Transit Administration. (2015). Reporting instructions for the Section 5309 - Capital investment grant program. Washington, DC: USDOT/FTA.

Federal Transit Administration. (2016). Standard Cost Categories (SCC) for Capital Projects. Retrieved from Federal Transit Administration: https://www.transit.dot.gov/funding/grant-programs/capital-investments/standard-cost-categories-scc-capital-projects

Federal Transit Administration. (2007). The predicted and actual impacts of new starts projects — 2007 -- capital cost and ridership. Washington, DC: USDOT/FTA.

Federal Transit Administration. (2016). Transportation Planning. Retrieved from Federal Transit Administration: www.transit.dot.gov/regulations-and-guidance/transportation-planning/transportation-planning

GAO. (2009). GAO Cost estimating and assessment guide: best practices for developing and managing capital program costs. United States Government Accountability Office.

Goodman, A. S., & Hastak, M. (2006). *Infrastructure planning handbook: Planning, engineering, and economics ; [environmental and social impact assessments, legal, institutional, and contractual aspects, planning for uncertainty and risk, domestic and international projects]*. New York, NY: McGraw-Hill Professional.

Grava, S. (2003). Urban transportation systems: choices for communities. McGraw-Hill.

Gunduz, M., Ugur, L. O., & Ozturk, E. (2011). Parametric cost estimation system for light rail transit and metro trackworks. *Expert Systems with Applications*, *38*(3), 2873–2877.

Harbuck, R. (2002). Using models in parametric estimating for transportation projects. 2002 AACE International Transactions, EST.05.1 - EST.05.9.

Harbuck, R. (2007). Are accurate estimates achievable during the planning of transportation projects? . 2007 AACE International Transactions, EST.16.1 - EST.16.6.

Harmatuck, D. (2008). Light rail cost functions and technical inefficiency. Journal of the Transportation Research Board, Vol 2042, 58 - 70.

Hoback, A. (2008). Sensitivity analysis of light rail transit unit capital costs. TRB 2008 Annual Meeting.

Hollmann, J. (2006). Total cost management framework: An integrated approach to portfolio, program, and project management. Morgantown, W. Va: AACE International.

Hsu, L. R. (2013). Cost estimating model for mode choice between light rail and bus rapid transit systems. Journal of Transportation Engineering, 20 - 29.

Indianapolis Transit Consultants (ITC) - Schimpeler/American, & Jakes Associates, Inc. (2003). *Technology Assessment (Tech Memo I-7A).*

Kendall, M. (1970). Rank correlation methods. London: Charels Griffin & Co. Ltd.

Kouskoulas, V. (1984). Cost functions for transportation systems. Journal of Transportation Engineering, Vol.110, No.1, 129 - 141.

Lattin, J. M. (2003). Analyzing Multivariate Data. Pacific Grove: Brooks/Cole – Thomas Learning.

Lehman, A., O'Rourke, N., Hatcher, L., & Stepanski, J. (2005). *Jmp(R) for basic Univariate and Multivariate statistics: A step-by-step guide*. Cary, NC: SAS Institute Inc.

Munns, A. (2000). Estimating using cost significant global models. Construction Management and Economics, 18, 575 - 585.

Newnan, D. G. (1983). Engineering economic analysis. San Jose, CA: Engineering Press.

Ostle, B. (1963). Statistics in research. The Iowa State University Press.

Paaswell, R. (2005). Analysis of capital cost elements and their effects on operating costs. Washington, DC: Federal Transit Administration.

Pickrell, D. (1990). Urban rail transit projects: forecast versus actual ridership and costs. Washington, DC: Urban Mass Transportation Administration.

Priemus, H. Flyvbjerg, B., & van Wee, B. (2008). • Decision-Making on Mega-Projects: Cost-Benefit Analysis, Planning and Innovation. Cheltenham: Edward Elgar Publishing Limited.

Siegel, S., & Castellan, N. (1988). Nonparametric statistics for the behavioral sciences. McGraw-Hill Book Co.

Thompson, B. (2004). Exploratory and confirmatory factor analysis: understanding concepts and applications. Washington, DC: American Psychological Association.

Touran, A. (1993). Probabilistic cost estimating with subjective correlations. *Journal of Construction Engineering and Management*, Vol. 119, No.1, 58 - 71.

Touran, A. (2010). Probabilistic approach for budgeting in portfolio of projects. *Journal of Construction Engineering and Management*, *136*(3), 361–366.

Transit New Zealand (1993). The reliability of capital cost estimates. Wellington, N.Z

Trost, S. M., & Oberlender, G. D. (2003). Predicting accuracy of early cost estimates using factor analysis and Multivariate regression. *Journal of Construction Engineering and Management*, *129*(2), 198–204.

USDOT. (2015 (update)). Public Involvement Techniques for Transportation Decisionmaking. USDOT/FHWA.

Van de Geer, J. (1971). Introduction to multivariate analysis for the social sciences. San Francisco: W.H. Freeman and Company.

Waie, P. (2008). RSMeans building construction cost data 2009. Kingston: R.S. Means Co., Inc.

Whiteside, J. (2004). Developing estimating models. Cost Engineering, Vol. 46, No. 9, 23 - 30.

Wilmot, C. G., & Cheng, G. (2003). Estimating future highway construction costs. *Journal of Construction Engineering and Management*, *129*(3), 272–279.

VITA

Anthony J. Catalina

EDUCATION

Ph.D, Civil Engineering
University of Kentucky, Anticipated 2016

M.S., Civil Engineering,
University of California-Irvine, 1990

B.S., Geography,
Michigan State University, 1985

PROFESSIONAL EXPERIENCE

Transit Division Manager, American Consulting Professionals, LLC, 2010 – 2016

*Transit Division Manage*r, ENTRAN, 2000 to 2010

*Associate/Project Manager/Sr. Transportation Planne*r, Carter & Burgess, 1998 to 2000
Associate Vice President, Frederic R. Harris, 1998

Vice President, Jordan, Jones & Goulding / Schimpeler Associates, 1995 to 1998

Executive Vice President, Chaiseri & Schimpeler Co., 1996 to 1998

*Vice Presiden*t, Mott MacDonald, Inc./ Schimpeler Associates, 1995 to 1996

Principal Engineer / Senior Planner, ICF Kaiser Engineers, 1992 to 1995

Associate / Transportation Planner, Schimpeler-Corradino Associates, 1987 to 1992

Transportation Planner, West Michigan Regional Planning Commission, 1986 to 1987

Transportation Planner, Tri-County Regional Planning Commission, 1985 to 1986

PUBLICATIONS

Stopher, P., Metcalf, H., Wilmot, C., ***Catalina, A***., & Schimpeler, C. (1999). Estimating patronage for a feasibility study of high-speed rail in Thailand. *Transportation Research Record: Journal of the Transportation Research Board*, *1691*, 12–18.